Continuation Fields: Waves, Admissibility, and the Selection of Thought

Flyxion
Independent Researcher

September 2026

Preface

A brain has a fixed set of connections at any given instant, and it also has a train of thought. Neither fact explains the other. The connections change too slowly to account for how quickly attention reorganizes, and the train of thought is too structured to be the product of connections alone. Something has to mediate between a stable repertoire of possible associations and the particular one that becomes active now, and whatever mediates has to operate on a timescale close to the thought itself rather than the timescale of synaptic change.

This book takes seriously a specific proposal about what that mediator is. Traveling electrical waves across the cortex, organized in frequency, phase, and spatial pattern, have been measured for close to a century,[1] but their function has remained genuinely contested. A 2026 review by Earl Miller, Scott Brincat, and Jefferson Roy [11] argues that these waves are not simply a byproduct of neural firing but an active organizing layer: synaptic connectivity supplies a space of possible representations, and wave dynamics determine which subset of that space is expressed, bound together, and carried forward at a given moment. The architecture can be stated compactly as a chain,

$$\mathcal{S} + q_t \longrightarrow z_t, \qquad q_t \supseteq \mathcal{A}_t^{\text{wave}} \quad \text{as the field hypothesis,}$$

where $\mathcal{S}$ is the relatively persistent space of synaptic dispositions, q_t is whatever fast, temporally varying quantity organizes the population's instantaneous state z_t, and $\mathcal{A}_t^{\text{wave}}$ names the specific, structured field this book investigates as a candidate contributor to q_t. The claim under examination is not that q_t replaces $\mathcal{S}$, or that wiring is unimportant, and it is not yet a claim that q_t must be realized as an oscillatory field rather than as ordinary recurrent network dynamics, a question the book takes up directly and does not presuppose. What is proposed, more narrowly, is that the same $\mathcal{S}$ supports different trajectories because some fast, temporally varying quantity organizes which trajectory is realized, and that structured wave dynamics are a measurable candidate contributor to this organization whose causal independence from recurrent state remains to be established.

This book is built around that one paper, read closely rather than surveyed broadly. It extends the argument where the underlying evidence supports extension, and it stops where the paper itself stops, since the authors are unusually direct about the limits of what they have shown. Two questions in particular are carried through every part of the book without being resolved. The first is whether wave interference is itself doing computational work, in the sense of transforming information through continuous physical interaction, or whether waves are better read as a correlate of computation performed elsewhere in recurrent synaptic circuitry. The second is whether the loss of an organized, integrated dynamical regime is part of the mechanism of consciousness or merely covaries with it. Both questions are treated as open at every point they arise, including in the final chapter, because closing them prematurely would misrepresent what the evidence currently

[1] Hans Berger's first published recordings of human cortical rhythms date to 1929 [4]. What has changed since is not the existence of the signal but the resolution at which its spatial and phase structure can be tracked across simultaneously recorded sites.

supports.

The chapters that follow use a small amount of formal vocabulary, developed from a broader project on admissibility, repair, and continuation as general operations on constrained possibility spaces. That vocabulary is introduced only after each empirical problem has been established on its own terms, so that the mathematics compresses a distinction already earned by the neuroscience rather than supplying the distinction in advance. For the same reason, the fifth and final part of the book, which gathers the formal apparatus into a single calculus, is deliberately placed last. Nothing in that part is a new claim. It is a restatement, in compact notation, of what the preceding four parts established or left open.

The result is a narrower claim than the one sometimes suggested by popular treatments of oscillatory binding, and a more defensible one than a theory of consciousness built directly on wave mechanics. The proposal argued for here is that cognition may require a continuously reconstructed field of admissibility, and that the geometry of that field determines which latent distinctions in a fixed representational repertoire can become a coherent continuation of thought right now. Whether that geometry is doing genuine computational work, and whether its collapse is what consciousness loses when it is lost, are questions this book raises with as much precision as it can and answers with as much honesty as it can manage, which in both cases falls short of certainty.

Part I

The Underdetermination Problem

Chapter 1

A Brain With Too Many Jobs Per Neuron

A convenient fiction about the brain, useful in introductory teaching and almost immediately misleading beyond it, is that neurons specialize. A face cell responds to faces. A place cell responds to a location. Extend that picture upward into the prefrontal cortex, where planning, rule-following, and working memory are thought to live, and it breaks down quickly. Neurons there routinely respond to combinations of task variables rather than to any single one, and the combination a given neuron cares about is often unstable from one context to the next. A cell that fires for a particular color may do so only under a particular rule, and only when a particular location is also being held in memory. Change the rule and the same cell's tuning to color may change or disappear entirely. This pattern has a name in the systems neuroscience literature: mixed selectivity.

The natural first reaction to mixed selectivity is that it looks like noise, or like evidence that single neurons are a poor unit of analysis for understanding cognition. Rigotti and colleagues showed, in a result that has held up and been extended substantially since [14], that the opposite is closer to the truth. A population of neurons with mixed selectivity can represent a far larger number of task-relevant distinctions than an equally sized population of cleanly tuned, single-purpose neurons. The reason is combinatorial. If each neuron is committed to one variable, the population's representational capacity grows additively with the number of neurons. If neurons instead respond to conjunctions and interactions among variables, capacity grows combinatorially, because a given neuron's response can participate in many different task-relevant distinctions depending on what else is currently true. High-dimensional, mixed-selectivity representations are what let a fixed number of neurons support flexible, open-ended behavior rather than a small fixed repertoire of stimulus-response pairings.

That capacity comes at a structural cost, and the cost is the subject of this book. If a neuron's functional identity is not fixed, but instead depends on the surrounding context of rule, goal, and other represented content, then something has to specify that context moment to moment. A neuron does not know, by virtue of its wiring alone, which of its many possible jobs it is currently doing. Two trials that activate the same population of prefrontal neurons can correspond to two entirely different computations, distinguished only by which task-relevant conjunction is currently being read out. Without some mechanism to settle that ambiguity on each trial, mixed selectivity would produce exactly the kind of crosstalk its critics originally worried about: high representational capacity purchased at the price of interpretability, both for downstream neural populations and for anyone trying to understand what the population is doing.

The obvious candidate for settling the ambiguity is the wiring itself: perhaps synaptic weights

specify, precisely enough, which conjunction of inputs drives which output in which context, and no further mechanism is needed. The wiring alone, understood as a static matrix of synaptic parameters, does not uniquely determine which functional assignment is realized at a particular moment, since the same weights are compatible with many different momentary outputs depending on the population's current state and current input. But this underdetermination does not mean that the wiring must be physically remodeled whenever the assignment changes. A recurrent network with entirely fixed weights can occupy rapidly changing states, moving among attractors, transient trajectories, and effective-connectivity regimes on a timescale far faster than synaptic plasticity, and those fast-changing states can by themselves alter how subsequent inputs are transformed. Fixed connectivity is therefore not obviously too slow to resolve the ambiguity. It may already contain, in its dynamics rather than in its structure, everything needed to do so.

The relevant contrast is accordingly not between slow synaptic plasticity and some faster alternative mechanism. It is between a relatively persistent parameter structure, the synaptic weights themselves, and the fast state variables that select a particular trajectory within the space those weights make possible. The open question is which fast variables are doing the selecting, and whether they are exhausted by the population's own recurrent state or require something further. Candidates include local and recurrent network state, short-term synaptic dynamics operating on the timescale of hundreds of milliseconds,[1] neuromodulatory context, traveling oscillatory field geometry, or some coupled system involving several of these together. Nothing said so far rules any of them out.

This is the precise shape of the problem the rest of the book is organized around. It is not the vague complaint that the brain is complicated or that wiring diagrams alone cannot capture everything interesting about the mind, and it is not yet a case for any particular fast variable over any other. It is a specific claim that persistent structure underdetermines momentary output, together with an open question about what supplies the rest. The next chapter states this requirement formally. The chapter after that takes seriously the strongest available answer to it that does not require anything beyond recurrent network dynamics, before the remainder of the book makes the case for a further ingredient.

[1] Short-term facilitation and depression can alter synaptic efficacy rapidly, while their onset and recovery constants vary from tens of milliseconds to seconds across synapse types. They therefore remain plausible contributors to q_t, not mechanisms excluded solely by timescale.

Chapter 2

Repertoire Is Not Selection

The claim that ended the previous chapter can be stated more precisely than "wiring cannot explain cognition," and it should be, because the imprecise version invites an easy and unhelpful rebuttal: of course wiring matters, nobody denies it, so what exactly is being claimed instead. The precise version is an underdetermination claim, and underdetermination claims are a familiar and respectable kind of argument. They do not say that a given structure is irrelevant. They say that the structure, on its own, does not fix the outcome in question, and that something further is needed to fix it. Getting the precise version right matters more than usual here, because an imprecise statement of it risks quietly ruling out the strongest competing account before that account has had a chance to answer.

A first pass might define $\mathcal{S}$ as the space of representational dispositions a population's synaptic connectivity makes available, and γ_t as the trajectory realized at time t, then write $\mathcal{S} \not\Rightarrow \gamma_t$ and treat the matter as settled. This is too fast. A fixed dynamical law, together with a complete specification of the system's current state and its full input history, can determine a trajectory perfectly well without requiring anything beyond the connectivity that implements the law. If $\mathcal{S} \not\Rightarrow \gamma_t$ is to say something nontrivial, $\mathcal{S}$ has to be understood narrowly, as connectivity and disposition space alone, excluding the current state and the ongoing stream of inputs and endogenous modulation that any dynamical account would also require. Once that narrowing is made explicit, the claim is true, but it is true for an unsurprising reason: no static structure, on its own and without a current state or an input history, determines what happens next. Making this explicit is what allows the claim to avoid being trivial in one direction and unfalsifiable in the other.

The fuller formalization makes the substantive content of the claim, and the room left for competing answers, easier to see. Let z_t denote the instantaneous population state at time t, the quantity actually evolving moment to moment, distinct from $\gamma_{[t_0,t_1]} = \{z_t : t \in [t_0, t_1]\}$, the trajectory traced out over some interval. This distinction is introduced now because Part III will need to compare a pre-distraction trajectory, a post-distraction trajectory, and a counterfactual uninterrupted trajectory, and conflating the instant with the interval would cause real trouble there. The state evolves according to

$$z_t = F_{\mathcal{S}}\big(z_{t_0}, u_{[t_0,t]}, \eta_{[t_0,t]}\big)\,,$$

where u is exogenous input and η collects endogenous modulation and noise, and $F_{\mathcal{S}}$ is the dynamical law implemented by the connectivity $\mathcal{S}$.[1] Given complete knowledge of $\mathcal{S}$, the initial state, and

[1] $F_{\mathcal{S}}$ is left deliberately unspecified as to functional form. It could be a rate model, a spiking network, or a mean-field approximation; nothing in the underdetermination argument depends on that choice, and committing to one prematurely would let a modeling convenience masquerade as a substantive claim about neural computation.

the full input and modulation history, z_t is determined. The underdetermination claim, restated properly, is not that $\mathcal{S}$ fails to determine z_t under those conditions. It is that $\mathcal{S}$ alone, without an accompanying account of what organizes the state and input history moment to moment, does not by itself explain how a population resolves the ambiguity described in the previous chapter, since z_{t_0}, u, and η are doing unspecified work in that equation, and specifying what that work consists of is precisely the open problem.

Call whatever fast, temporally varying quantity is doing that organizing work q_t, without yet committing to what q_t is made of:

$$\mathcal{S} + q_t \longrightarrow z_t.$$

This is deliberately the most permissive formulation available. q_t could be exhausted by the population's own recurrent state, in which case Chapter 3's competitor is correct and nothing beyond ordinary recurrent dynamics is needed. It could include short-term synaptic facilitation and depression, neuromodulatory tone, or traveling field geometry, singly or in combination. The claim this book defends is the substantive hypothesis that q_t is not exhausted by recurrent state alone, and that a structured admissibility field, written $\mathcal{A}_t^{\text{wave}}$, contributes something to q_t that recurrent dynamics do not already provide:

$$q_t \supseteq \mathcal{A}_t^{\text{wave}}.$$

Once this hypothesis is granted for the sake of argument, the dynamics can be written with the field entering as a structured modulating variable rather than as an independent cause acting alongside z_t and u_t,

$$\dot{z}_t = F_{\mathcal{S}}(z_t, u_t; \mathcal{A}_t^{\text{wave}}).$$

This gives the hypothesis a precise burden rather than a rhetorical one. It must be shown either that $\mathcal{A}_t^{\text{wave}}$ improves prediction of z_t beyond what is available from z_t's own recent history, current input, and recurrent parameters, or that direct intervention on $\mathcal{A}_t^{\text{wave}}$, meaning a manipulation of wave geometry specifically, changes the resulting trajectory while the competing variables are held fixed.

Stating the hypothesis this way also guards against a specific failure mode. If q_t is simply defined as whatever fast variable turns out to be missing from a given account, then $\mathcal{A}_t^{\text{wave}}$ risks becoming a name attached after the fact to any residual variance a simpler model fails to explain, which would make the hypothesis unfalsifiable rather than merely difficult to test. The distinction between the generic placeholder q_t and the specific, independently measurable quantity $\mathcal{A}_t^{\text{wave}}$ is what prevents that collapse. q_t marks the shape of the open question. $\mathcal{A}_t^{\text{wave}}$ is this book's answer to it, and the answer has to earn its place against q_t's other candidates rather than being installed into the notation as though no other candidate existed. The next chapter takes the strongest of those candidates, recurrent network dynamics on their own, and gives it the direct hearing it is owed before the remainder of the book argues for extending q_t beyond it.

Chapter 3

Competing Accounts: Recurrence Without Fields

The previous chapter left open what fills the generic role of q_t, the fast, temporally varying quantity that, together with the persistent connectivity $\mathcal{S}$, determines the realized state z_t. It named a specific hypothesis, that q_t includes a structured admissibility field $\mathcal{A}_t^{\text{wave}}$, but it deliberately did not defend that hypothesis against its strongest rival. A substantial body of work in systems neuroscience answers the question of what fills q_t differently, and that work deserves a direct hearing here, both because it is empirically serious and because the case for including $\mathcal{A}_t^{\text{wave}}$ in q_t is only as strong as its margin over this alternative.

The recurrent-network family of models treats a cortical population as a dynamical system whose state evolves according to its own internal connectivity together with its current inputs. On this view, q_t is exhausted by the population's own recurrent state: the pattern of activity currently circulating through the network's recurrent connections, which shapes how the network responds to the next input and therefore which of the many dispositions in $\mathcal{S}$ becomes expressed. No additional field variable is needed, on this account, because the recurrent dynamics already constitute a temporally varying selection mechanism sufficient to fill q_t on its own. A network can implement context-dependent computation, switch between effectively different input-output mappings depending on its recent history, and recover from a transient perturbation back toward an attractor state, all using ordinary recurrent connectivity and ordinary spiking, without invoking oscillatory fields, traveling waves, or ephaptic coupling as a separate causal layer.

This is not a weak alternative. Dynamic effective-connectivity models, in which the effective coupling between neural populations is treated as itself time-varying and estimated from data,[1] can account for a wide range of task-dependent reconfiguration phenomena, including some of the same phenomena the wave-based literature this book draws on describes: shifting patterns of which regions influence which other regions, context-dependent gating of information flow, and recovery of task-relevant state after interruption. Proponents of this family of models can point out, correctly, that recurrent dynamics and oscillatory dynamics are not independent phenomena to begin with. Local field potentials, the signal from which most of the oscillatory evidence in this book is drawn, are themselves generated by synchronized synaptic and spiking activity in local populations. It is possible to read the entire wave-based literature as a description, at a coarser measurement scale,

[1] Dynamic causal modeling, developed principally by Friston and colleagues [8], is the most extensively worked-out framework of this kind, and has been applied to exactly the sort of task-dependent reconfiguration this chapter describes. It is cited here as a representative methodology, not because this book adopts its specific generative-model formalism.

of the same recurrent computation that a spiking-network model would describe at a finer scale, in which case talk of an admissibility field is redundant with talk of recurrent state, differing only in level of description rather than in what is actually doing the causal work.

If that reading is correct, this book's project would be a relabeling exercise: taking a phenomenon adequately described by recurrent network theory and re-describing it in the vocabulary of fields, admissibility, and continuation, without adding anything the recurrent account did not already provide. That would not make the redescription false, but it would make it explanatorily idle, and a redescription that adds no explanatory or predictive purchase is not worth the additional formal apparatus it requires.

The condition under which the field-based account is not idle can therefore be stated precisely, and it is worth stating before any of the supporting evidence is presented, so that the evidence can be judged against a standard set in advance rather than a standard adjusted after the fact. The field account earns its distinctiveness only if some property of the wave itself, specifically its spatial phase, its direction and speed of propagation, or the geometry of its interference with other waves, carries explanatory or causal information about γ_t that is not reducible to a description of local spiking rates and recurrent synaptic interaction. If oscillatory measures track behavior only insofar as they track underlying recurrent state, and add nothing once recurrent state is accounted for, the recurrent account is preferable on grounds of parsimony, and this book's more elaborate formal apparatus should be set aside in favor of it.

This condition is not asserted lightly, and it is not treated as a formality to be satisfied once and then forgotten. It recurs explicitly in Chapter 8, where the decisive experiment capable of adjudicating it is described in full, and again in Chapters 11 and 16, where the recovery and consciousness evidence respectively are checked against it directly rather than simply added to a growing list of correlations. The chapters that follow present evidence that is, at minimum, consistent with the field account clearing this bar. Whether it is sufficient to clear it is a question this book raises honestly rather than answers by assumption, and the reader is owed the reminder, repeated at each of those later points, that the recurrent-network account given here has not been ruled out.

Part II

Admissibility as a Relational Dynamic

Chapter 4

Against Fixed Roles for Frequency Bands

Part I established that some fast, temporally varying quantity q_t must accompany the persistent connectivity $\mathcal{S}$ to determine a population's realized state, and named the hypothesis that a structured admissibility field, $\mathcal{A}_t^{\text{wave}}$, is at least part of what fills that role. This part turns to the evidence for that hypothesis, beginning with the specific pattern most often cited in its favor: a rough division of labor between slower and faster oscillatory rhythms, in which alpha and beta activity, in the range of roughly ten to thirty cycles per second,[1] appear to carry information about task rules and goals, while gamma-range activity appears more closely tied to the moment-to-moment expression of specific held content. Lundqvist and colleagues, recording from prefrontal cortex during working-memory tasks [9], found gamma bursts co-occurring with the encoding and recall of specific items, while beta bursts tracked something closer to a default, rule-governed state that gamma activity had to break through in order to express content.

This finding is worth taking seriously and worth being careful with in roughly equal measure. It would be a mistake to read it as establishing that alpha and beta activity simply are the admissibility field, and that gamma bursts simply are content expression, as though frequency band and function were the same thing observed at two different resolutions. Frequency-functional associations of this kind have not held up as fixed assignments elsewhere in the literature. The same nominal frequency band can carry different information, or play a different functional role, depending on the cortical area recorded from, the species studied, the task demands in place, and even the specific method used to define the band's boundaries. A rhythm that looks like rule-maintenance in one recording context can look like something closer to stimulus-driven sensory processing in another. Treating "alpha and beta constrain, gamma expresses" as a general law of cortical function would overstate what a specific and well-controlled set of working-memory experiments has shown.

The more defensible claim is relational rather than definitional, and it is the claim this book adopts going forward. What the evidence supports is a relationship between two timescales of oscillatory activity, in which relatively slow, spatially extended dynamics constrain which of a population's faster, more spatially localized, content-bearing events are currently permitted to occur. This relationship can be instantiated by alpha and beta rhythms constraining gamma bursts in the specific experimental paradigms Lundqvist and colleagues studied, without committing the

[1] These boundaries are conventional rather than physiologically sharp. Different laboratories draw the alpha/beta and beta/gamma divisions at somewhat different frequencies, and the same recording can be partitioned differently depending on the spectral estimation method used. This is part of why the chapter goes on to insist that the functional claim be kept separate from any particular numerical cutoff.

theory to the claim that alpha and beta are, as a matter of definition, always and everywhere the constraining rhythm, or that gamma is always and everywhere the constrained content. In a different task, a different cortical area, or a different species, the same relational role might be filled by a different pair of frequency bands, or by a relationship that does not decompose into two clean bands at all.

This distinction matters for how the formal apparatus of the next chapter should be read. The operators introduced there, drawn from a broader vocabulary this book borrows for naming transitions between admissible and inadmissible states, are defined independently of any specific frequency band. A transition into current expression is what the framework calls a certain kind of event, regardless of which oscillatory signature, if any, turns out to implement it in a given recording. Gamma bursting is then a candidate marker of that kind of event in the specific paradigms where it has been measured, available to be confirmed, qualified, or overturned by further evidence, rather than a definitional stand-in for the event itself. Keeping the operator and its candidate neural marker separate is what allows the theory to be revised by future data about which frequencies do what, without having to abandon the underlying claim about admissibility and expression each time a frequency-specific finding is refined or reversed.

One further caution belongs here before the formal chapter that follows. The evidence reviewed so far establishes, at most, that oscillatory measures are informative about task structure and content, in the sense that reading them off a recording lets an observer predict rule state or item identity above chance. This is the first and best-supported tier of the three-part hierarchy this book uses throughout: waves as informative measurements of an underlying organization, whatever that organization turns out to be made of. It does not yet establish that the oscillation is causally regulating which content becomes expressed, which is the second, more demanding tier, and it certainly does not establish that wave interference is itself performing a computation not reducible to recurrent synaptic dynamics, which is the third and most demanding tier and remains, at this point in the book, entirely open. The chapters that follow move through evidence bearing on the second tier before confronting the third directly.

Chapter 5

Formalizing POP, REFUSE, BIND, and COLLAPSE

The preceding chapter separated an operation from any particular neural marker proposed to implement it. That separation is methodologically necessary. If POP is defined as gamma activity, then evidence that gamma serves a different function in another cortical region would refute the definition before it tested the theory. If POP is instead defined as a transition by which a latent distinction becomes presently expressed, gamma bursting can be evaluated as one candidate realization of that transition. The operator remains stable while the empirical assignment remains corrigible.

This chapter defines four such operations: POP, REFUSE, BIND, and COLLAPSE. They are not introduced as elementary particles of cognition, nor as four newly discovered neural mechanisms. They are substrate-neutral descriptions of changes that can occur within a constrained possibility space. Their value depends upon whether they divide neural dynamics into empirically distinguishable classes and whether candidate measurements can fail to satisfy their defining conditions. An operator that applies to every possible observation explains none of them.

The construction begins with four sets and two timescales. Let $\mathcal{S}$ denote the relatively persistent disposition structure established by connectivity, learned representations, and other slowly changing parameters. Let

$$P(\mathcal{S})$$

denote the set of distinctions and continuations that this structure can support under some attainable condition. Membership in $P(\mathcal{S})$ is therefore dispositional rather than actual. An item may belong to this space while remaining neither currently accessible nor currently expressed.

Let $A_t \subseteq P(\mathcal{S})$ denote the distinctions admissible at time t. Admissibility means more than physical possibility and less than actual expression. An admissible distinction is available to influence the current computation, enter working expression, constrain an action, or participate in a continuing trajectory. Let

$$E_t \subseteq A_t$$

denote the distinctions presently expressed in measurable neural activity or behavior. Finally, let z_t denote the instantaneous neural state and

$$\gamma_{[t_0,t_1]} = \{z_t : t \in [t_0, t_1]\}$$

the realized trajectory over an interval.

The resulting inclusions are

$$E_t \subseteq A_t \subseteq P(\mathcal{S}).$$

They distinguish what the system is expressing, what it could presently express, and what its persistent organization could express under some other attainable condition. These levels should not be identified. A representation can remain in $P(\mathcal{S})$ after it has become inadmissible, and it can remain admissible without being expressed. Much of the theory developed in this book concerns transformations among these statuses.

The notation does not assume that A_t is already known to be a traveling electrical field. At this stage, A_t is an operationally defined restriction on effective possibility, playing the role Chapter 2 assigned to the generic organizing quantity q_t: recurrent state, neuromodulatory context, oscillatory phase, traveling-wave geometry, and external task structure may all contribute to it. The field hypothesis developed later makes the stronger claim that spatially organized wave dynamics constitute a causally important part of this restriction, in the notation of Chapter 2 that $q_t \supseteq \mathcal{A}_t^{\text{wave}}$. The definitions below do not receive that claim for free.

POP: Entry Into Expression

Let x be a distinction supported by the system but not presently expressed. A POP is a transition by which x enters the expressed set:

$$\text{POP}_{t,\Delta t}(x) : x \in A_t \setminus E_t \longmapsto x \in E_{t+\Delta t}.$$

The definition requires prior admissibility. An event that forces a wholly unavailable distinction into the system by changing its disposition structure is not a POP in this sense; it is learning, construction, or structural modification. POP operates within a currently available repertoire.

Expression must be defined relative to an observation scale. A remembered color might be expressed in a decodable population pattern without producing overt behavior. A motor intention might be expressed in premotor cortex while remaining absent from muscle activity. We may therefore write

$$E_t^{(\Omega)}$$

for expression relative to an observation map Ω. A neural POP and a behavioral POP need not occur simultaneously:

$$\text{POP}^{(\text{neural})}(x) \not\Rightarrow \text{POP}^{(\text{behavioral})}(x).$$

This prevents lack of overt report from being treated automatically as lack of neural expression.

A candidate neural marker M_t realizes POP only if changes in that marker predict or cause the transition of a specific distinction from admissibility into expression. Mere temporal coincidence is insufficient. If gamma bursts are proposed as a marker, the relevant claim is not simply

$$M_t^{\gamma} > 0.$$

It is that variation in the occurrence, timing, location, or structure of the gamma event bears a reliable relation to the entry of x into E_t:

$$\Pr(x \in E_{t+\Delta t} \mid M_t^{\gamma},\, x \in A_t \setminus E_t) > \Pr(x \in E_{t+\Delta t} \mid \neg M_t^{\gamma},\, x \in A_t \setminus E_t)\,.$$

For a causal interpretation, intervention on the marker must alter the transition while appropriate background conditions are controlled.

The definition can fail in several ways. A gamma burst may occur without any identifiable content entering expression. Content may enter expression consistently without gamma bursting. Gamma may reflect maintenance, routing, error correction, or motor preparation rather than entry into expression. Any of these findings would weaken the proposed gamma-POP correspondence without eliminating POP as a formal category.

POP is therefore an event of actualization, but it is not yet a claim about conscious access. A distinction may enter an operative neural expression without becoming reportable, globally available, or phenomenally experienced. Those stronger transitions require additional conditions that the present definition deliberately omits.

REFUSE: Active Exclusion From Admissibility

REFUSE is more demanding to define because absence is cheap. At any moment, nearly every distinction in $P(\mathcal{S})$ is absent from expression. If all such absence counted as refusal, REFUSE would name the generic condition of not happening rather than a specific operation.

The first necessary condition is that the refused distinction remain dispositionally possible:

$$x \in P(\mathcal{S}).$$

The second is that it be excluded from present admissibility:

$$x \notin A_t.$$

These conditions alone are still insufficient. A permanently irrelevant memory, an unstimulated sensory feature, and an impossible response under the current task would all qualify. REFUSE requires a counterfactual contrast: under an appropriately matched condition lacking the candidate excluding process, x would have had a non-negligible probability of becoming admissible or expressed.

Let C_t denote the observed context and let C_t^{-r} denote a matched counterfactual context in which the candidate refusing process r is absent while relevant background conditions are preserved. Then r counts as a REFUSE operation on x only if

$$x \in P(\mathcal{S}), \qquad x \notin A_t,$$

and

$$\Pr\big(x \in A_{t+\Delta t} \mid C_t^{-r}\big) - \Pr(x \in A_{t+\Delta t} \mid C_t) > \varepsilon$$

for some experimentally justified threshold $\varepsilon > 0$.

Equivalently, one may define the refusal strength

$$\rho_r(x,t) = \Pr\big(x \in A_{t+\Delta t} \mid C_t^{-r}\big) - \Pr(x \in A_{t+\Delta t} \mid C_t)\,.$$

Positive ρ_r indicates exclusion attributable to r; $\rho_r \approx 0$ indicates mere non-occurrence; negative ρ_r indicates that the supposed refusing process facilitates rather than suppresses admissibility.

This counterfactual condition distinguishes refusal from simple inactivity. Suppose a distractor representation is decodable at the sensory periphery but prevented from influencing a working-memory decision. If perturbing a spatial alpha/beta pattern allows the distractor to enter the task-relevant population state or alter behavior, the original pattern is evidence for REFUSE. If removing the pattern changes nothing, then its coexistence with distractor suppression does not establish that it performed the suppression.

REFUSE is also distinct from erasure. Its characteristic structure is

$$x \in P(\mathcal{S}) \quad \text{and} \quad x \notin A_t.$$

The distinction remains supported by the persistent disposition space even while current organization excludes it. This makes refusal reversible in principle: a later change in context can restore x to admissibility without rebuilding the representation. Structural forgetting, by contrast, would alter $P(\mathcal{S})$ itself.

The counterfactual definition creates a demanding empirical burden. Counterfactual neural conditions cannot simply be observed; they must be approximated through intervention, matched trials, causal modeling, or carefully justified natural variation. That difficulty is a virtue rather than a defect. It prevents every region of high alpha/beta power from being declared a refusing boundary merely because task-relevant spiking happens to be weak there.

BIND: Formation of a Jointly Operative Distinction

Two expressed contents can coexist without being bound. A population may carry information about a color and a location while failing to represent which color occurred at which location. Temporal coincidence, anatomical proximity, and phase synchronization are therefore possible conditions or mechanisms of binding, not definitions of binding itself.

Let x and y be admissible distinctions. BIND constructs a joint distinction

$$\langle x, y \rangle_\kappa$$

whose organization depends upon a relation κ. The relation might be temporal order, spatial correspondence, role assignment, phase relation, common task membership, or another empirically specified structure:

$$\mathrm{BIND}_\kappa(x, y) : (x, y) \longmapsto \langle x, y \rangle_\kappa.$$

The result is not merely the set $\{x, y\}$. It preserves information about how the components belong together.

A minimal behavioral test of binding asks whether the system can discriminate the correct conjunction from recombinations containing the same components. If red appeared on the left and blue on the right, a bound representation must distinguish

$$\langle \text{red}, \text{left} \rangle$$

from

$$\langle \text{red}, \text{right} \rangle,$$

even though both alternatives contain representations of red and spatial position. Successful component decoding alone does not establish binding.

An information-theoretic criterion can state this requirement more generally. Let T be a task variable whose correct determination depends upon the relation between x and y. Let X and Y denote measurements associated with the component representations, and let J_κ denote a measurement of their proposed joint organization. Evidence for binding requires that J_κ carry task-relevant information not exhausted by the component measurements considered separately:

$$I(J_\kappa; T) > I(X; T) + I(Y; T) - R(X, Y; T) + \varepsilon,$$

where $R(X, Y; T)$ corrects for redundant information shared by X and Y, and $\varepsilon > 0$ is an evidential threshold.

Because redundancy and synergy cannot be recovered reliably from the simplest mutual-information expression alone, a partial-information decomposition is preferable where data permit it:[1]

$$I(X, Y; T) = R + U_X + U_Y + \mathrm{Syn}(X, Y; T).$$

Here R is redundant information, U_X and U_Y are the components' unique contributions, and

$$\mathrm{Syn}(X, Y; T) > 0$$

indicates information available from the joint state that is unavailable from either component independently. Positive synergy is not, by itself, sufficient to prove a particular neural binding mechanism, but it supplies a principled criterion for the existence of a jointly informative organization.

When phase relation is the proposed binding variable, the stronger test is whether task-relevant conjunction information depends upon that relation:

$$I(J_\phi; T \mid X, Y) > 0.$$

A perturbation that shifts ϕ while approximately preserving component-level activity should then alter conjunction-sensitive decoding or behavior:

$$\mathrm{do}(\phi \mapsto \phi') \quad \Rightarrow \quad \Delta I(J_\phi; T) \neq 0$$

or

$$\mathrm{do}(\phi \mapsto \phi') \quad \Rightarrow \quad \Delta \Pr(\widehat{T} = T) \neq 0.$$

This is the form in which BIND connects to the field hypothesis. Phase alignment is not declared to be binding. It becomes evidence for binding only when manipulating or measuring phase explains the preservation of a joint distinction beyond what is explained by the component activities alone.

BIND may be transient. The joint distinction can dissolve while x and y remain individually admissible:

$$\langle x, y \rangle_\kappa \notin A_{t+\Delta t}, \qquad x, y \in A_{t+\Delta t}.$$

This is particularly important for distraction and anesthesia. A system can retain local content while losing the relations that make those contents part of one continuing computation.

COLLAPSE: Resolution Under Constraint

COLLAPSE is the most easily misunderstood of the four operations. In ordinary language, collapse suggests breakdown, destruction, or loss of organization. Within the present calculus it will mean something more specific: the resolution of multiple admissible continuations into a realized continuation under operative constraints. The failure of a system to sustain an organized regime will be called rupture, disintegration, or continuation failure, not COLLAPSE.

Let

$$\Gamma(A_t) = \{\gamma^{(1)}, \gamma^{(2)}, \ldots\}$$

[1] The decomposition is due to Williams and Beer [15]. Its specific numerical values depend on which of several proposed redundancy measures is used, since the decomposition into R, U_X, U_Y, and Syn is not uniquely fixed by the joint distribution alone. The qualitative criterion this chapter relies on, that synergy be reliably positive, is more robust to that choice than any particular decomposed value would be.

denote the continuations compatible with the current admissibility structure. COLLAPSE maps this constrained family to the trajectory that becomes realized:

$$\text{COLLAPSE}_{t,\Delta t} : \Gamma(A_t) \longmapsto \gamma^*_{[t,t+\Delta t]}.$$

The operation does not imply quantum measurement, metaphysical indeterminism, or irreversible destruction of unrealized alternatives. It says only that a system whose current organization admits several distinguishable continuations proceeds along one realized path.

If the dynamics are deterministic, the realized path may be fixed by the complete microstate even though it remains underdetermined relative to the coarser description A_t. If the dynamics are stochastic, collapse may be represented by a conditional distribution

$$\gamma^* \sim p(\gamma \mid A_t, z_t, u_t)\,.$$

The operator is compatible with either interpretation. Its explanatory role is relative to the level at which multiple continuations remain live.

COLLAPSE differs from POP. POP concerns the entry of a distinction into expression. COLLAPSE concerns selection among possible continuations. A POP may occur within a trajectory without selecting the trajectory as a whole, and a trajectory may collapse toward one action without introducing a newly expressed content. They frequently coincide in experiments because selecting a response often requires bringing response-specific information into expression, but the operations remain analytically distinct.

COLLAPSE also differs from REFUSE. REFUSE modifies the admissible set by excluding a possibility:

$$A_t \longmapsto A_t \setminus \{x\}.$$

COLLAPSE resolves among the continuations that remain:

$$\Gamma(A_t) \longmapsto \gamma^*.$$

Sequentially, refusal can shape collapse:

$$P(\mathcal{S}) \xrightarrow{\text{REFUSE}} A_t \xrightarrow{\text{COLLAPSE}} \gamma^*.$$

But collapse need not identify a single prior act of refusal. A continuous dynamical system can progressively concentrate around one trajectory as alternatives become unstable, inaccessible, or behaviorally irrelevant.

The empirical burden is to show that multiple continuations were genuinely admissible at the chosen level of description and that a measurable transition reduced this effective multiplicity. One possible quantity is the conditional entropy of future trajectories:

$$H(\Gamma_{t+\Delta t} \mid z_{\le t})\,.$$

A collapse-like transition should produce a marked reduction,

$$\Delta H_{\text{future}} = H(\Gamma_{t+\Delta t} \mid z_{\le t}) - H(\Gamma_{t+2\Delta t} \mid z_{\le t+\Delta t}) > 0,$$

provided that the reduction cannot be attributed merely to the passage of time or to additional sensory information. In practice, this may be estimated through state-space models, choice probabilities, trajectory decoding, or perturbational sensitivity.

A candidate oscillatory marker of COLLAPSE would therefore have to predict or control a reduction in effective future multiplicity. A transient increase in synchrony, for example, would not count simply because it looks globally organized. It would count only if it marked the point at which previously distinguishable continuations ceased to remain dynamically available and one task-relevant trajectory became reliably realized.

Composition of the Operators

The four operators describe different transformations, but cognition rarely presents them in isolation. Consider a working-memory task in which an animal must remember one of several objects, ignore a distractor, preserve the object's relation to a location, and choose a response after a delay. The target enters expression through POP. The distractor is prevented from entering the currently operative set through REFUSE. Object and location become jointly usable through BIND. The response period ends with COLLAPSE from several admissible actions to one realized action.

A simplified composition is

$$\mathcal{S} \xrightarrow{\text{REFUSE}} A_t \xrightarrow{\text{POP}} E_t \xrightarrow{\text{BIND}} E_t^{\text{joint}} \xrightarrow{\text{COLLAPSE}} \gamma^*.$$

This ordering is illustrative, not universal. Binding may alter what becomes admissible; a POP may change the boundary that governs later refusal; collapse may occur repeatedly at nested temporal scales. The operators are not stages in a mandatory pipeline. They form a vocabulary for identifying transformations within an evolving system.

Their interaction can be represented by a time-dependent operator family

$$\mathcal{O}_t = \{\text{POP}_t, \text{REFUSE}_t, \text{BIND}_t, \text{COLLAPSE}_t\},$$

acting upon the structured state

$$\Xi_t = (P(\mathcal{S}), A_t, E_t, z_t, \Gamma_t).$$

The evolution of the cognitive system can then be written abstractly as

$$\Xi_{t+\Delta t} = \mathcal{F}(\Xi_t, \mathcal{O}_t, u_t),$$

without yet specifying whether the operators are implemented by recurrent dynamics, traveling fields, neuromodulation, or their interaction.

This abstraction has a strategic purpose. It permits two theories to agree that REFUSE or BIND occurred while disagreeing about the physical mechanism. A recurrent-network account may identify REFUSE with attractor-dependent suppression and BIND with a conjunctive population code. A continuation-field account may identify REFUSE with a spatially patterned reduction of excitability and BIND with phase-dependent coordination across cortical locations. The theories can then be compared using common operational targets rather than talking past one another in mechanism-specific vocabularies.

The Three-Tier Evidential Test

Chapter 4 introduced three levels at which oscillatory dynamics may relate to these operations. The distinction can now be stated for each operator.

At the first level, a wave property is an informative measurement. It predicts that an operation has occurred:

$$I(M^{\text{wave}}; \mathcal{O} \mid C) > 0,$$

where C contains relevant controls. This establishes that the wave carries information about POP, REFUSE, BIND, or COLLAPSE. It does not establish that the wave performs the operation.

At the second level, the wave property causally regulates the operation:

$$p(\mathcal{O} \mid \text{do}(M^{\text{wave}} = m_1), C) \neq p(\mathcal{O} \mid \text{do}(M^{\text{wave}} = m_2), C).$$

Evidence at this level requires intervention or a comparably strong causal design. It shows that changing the wave changes the probability, timing, or structure of the operation.

At the third level, wave geometry contributes computational organization not captured by a competing description based on local spiking and recurrent state. Let G_t denote the relevant propagation or interference geometry and let R_t summarize the strongest available recurrent-state account. Irreducible contribution requires something like

$$I(\mathcal{O}; G_t \mid R_t, u_t) > 0,$$

together with intervention on G_t and evidence that the effect cannot be explained by uncontrolled changes in R_t. Even this conditional-information criterion is not sufficient by itself, because an incomplete measure of recurrent state can make a redundant field variable appear irreducible. The third tier therefore requires convergence between predictive improvement, geometry-selective intervention, and exclusion of plausible recurrent mediators.

These tiers prevent evidence from migrating upward without argument. A decodable phase pattern establishes the first tier. Ephaptic influence may help establish the second. Neither alone establishes the third. The claim that interference itself computes remains open until the geometry of interference is shown to transform task-relevant relations in a way not adequately captured by a model lacking that geometry.

What the Formalization Has and Has Not Achieved

The four operators now have meanings independent of any frequency band or recording technology. POP is entry into present expression. REFUSE is counterfactually supported exclusion from current admissibility. BIND is construction of a jointly operative distinction whose task-relevant information is not exhausted by its components. COLLAPSE is resolution from an admissible family of continuations to a realized trajectory. Each definition specifies observations that would fail to instantiate it.

Nothing in these definitions shows that the brain literally executes four discrete commands. Neural processes overlap, unfold continuously, and may admit several equally valid descriptive partitions. The operators identify functional transformations, not necessarily temporally isolated biological modules. Their legitimacy will depend upon whether they allow experiments to state more exact hypotheses than less structured language would permit.

Nor has the chapter established that A_t is an electromagnetic field in the physical sense. "Field of admissibility" remains, for the moment, a mathematical description of a spatially and temporally varying restriction on effective possibility. The next chapter turns to the evidence that makes the stronger interpretation plausible: cortical oscillatory activity forms structured spatial patterns, those patterns change with task demands, and their organization predicts trial-by-trial success and failure. That evidence does not yet show irreducible interference-computation. It does, however, move the argument from a formal possibility to a measured cortical geometry.

Chapter 6

Cortex as a Surface: Spatial Computing

The formal apparatus of the previous chapter converts a question that could otherwise drift indefinitely, whether waves organize cognition, into a narrower and more answerable one for any given body of evidence: which tier of the evidential hierarchy does this evidence reach, and for which operator. This chapter takes up the strongest available candidate for moving REFUSE and BIND beyond tier 1, the finding that alpha and beta activity is not simply a temporal rhythm but forms spatially organized patches across the cortical surface, and asks the narrow question honestly rather than assuming the answer in advance.

The findings themselves are worth stating plainly before any operator is attached to them. Lundqvist and colleagues, in a 2023 study [10], and Chen and colleagues, in a 2026 study [5], recording spikes and local field potentials from prefrontal cortex during working-memory and categorization tasks, found that alpha and beta power varies not only in time but across the physical extent of cortex, forming shifting spatial patches whose boundaries and locations change with task demands. Where a patch of strong alpha or beta power sits, spiking activity carrying task-relevant content tends to be suppressed; where the patch is weak or absent, such spiking is comparatively free to occur. Critically, the spatial organization of these patches predicts, on a trial-by-trial basis, whether the animal performs correctly or makes an error. This is measured, mesoscopic structure, not a redescription of an already-established temporal finding in spatial language.[1]

Consider first what this evidence does for REFUSE. The natural analysis available in this literature contrasts correct trials against error trials, asking whether a candidate suppressive patch was present, absent, or displaced on the trials where a distractor or an irrelevant response intruded. It is tempting to read this contrast as an approximation to the counterfactual condition Chapter 5's definition of REFUSE requires, comparing a context C_t in which the candidate refusing process operated against a context C_t^{-r} in which it did not. That reading should be resisted. Correct and error trials are not exchangeable with respect to the confounders that matter here. An error trial may differ from a correct one in arousal, sustained attention, the fidelity of sensory encoding, the state of motor preparation, or some other latent feature of network state that has nothing directly to do with the presence or absence of the candidate refusing patch. A trial-type contrast of this kind can identify a candidate refusing process and can estimate a predictive, refusal-like statistical

[1] "Spatial" here means spatial across the cortical surface as sampled by a multi-electrode array, typically on the order of millimeters between recording sites. This is coarser than the spatial scale at which individual dendritic fields or microcolumns operate, and the patch boundaries described in this literature should be understood at the array's resolution, not as a claim about finer-grained cortical geometry.

structure. It cannot, on its own, establish the causal quantity $\rho_r(x,t)$ that Chapter 5 defines refusal in terms of. What the correct-versus-error contrast is good for is motivating the intervention the formal definition actually demands, not standing in for that intervention as a weaker but acceptable substitute.

The case for BIND requires a comparably careful reading, and it is not automatically the stronger of the two simply because spatial organization predicts behavior. Trial-by-trial predictive power of the kind these studies report establishes, at most, that a spatial measure carries task-relevant information, $I(M^{\text{spatial}}; T \mid C) > 0$, which is exactly the tier-1 claim and nothing more. Chapter 5's criterion for BIND is specifically relational: it asks whether a proposed joint organization, across patches or across frequencies, carries information about a task variable that is not already available from its components considered separately, formally $\text{Syn}(X, Y; T) > 0$, or, where the proposed binding variable is phase, $I(J_\phi; T \mid X, Y) > 0$. Where the published spatial-computing analyses report that overall spatial or cross-frequency organization predicts behavior, without performing a partial-information decomposition or an equivalent conditional comparison against the components taken separately, the honest conclusion is narrower than "BIND has been demonstrated." The evidence identifies a plausible substrate on which binding could be implemented. It does not yet show that the specific joint quantity these studies measure carries information beyond what its components already provide. This is not a weakness in the calculus. It is the calculus doing exactly the discriminatory work it was built to do, refusing to let "structured and predictive" be quietly rewritten as "bound."

The tier assignment that follows from taking both corrections seriously is therefore more conservative than a first reading of the papers might suggest, and more informative for being so. Spatial patterning of alpha and beta power solidly clears tier 1 as a measure of task-state organization: it is decodable, it varies systematically with task demands, and it is genuinely spatial rather than merely temporal. It supplies a candidate tier-1 marker for both REFUSE and BIND. It does not, on the evidence considered so far, establish either operator under the book's own formal criterion, since REFUSE requires content-specific counterfactual exclusion that a correct-versus-error contrast cannot supply, and BIND requires a demonstration of relationally unique information that a bare predictive correlation does not provide. Trial-by-trial error prediction strengthens the association between spatial organization and task outcome. It does not, by itself, move either operator into tier 2, because tier 2 was defined in the previous chapter specifically in terms of intervention or a comparably strong causal design, and an observational contrast between naturally occurring trial types is neither.

What a genuine tier-2 experiment would need to look like can now be stated precisely, in the same idiom used to define the operators themselves. For REFUSE, the relevant manipulation would alter the geometry of a candidate suppressive patch directly, denoted $\text{do}(G_t = g')$, while holding total spectral power, average firing rate, sensory input, and task demands as fixed as the preparation allows. If the patch is doing the excluding, weakening or relocating it should selectively raise the admissibility of the previously excluded content, $\text{do}(G_t = g') \Rightarrow \rho_r(x,t) \downarrow$, accompanied by a measurable increase in content-specific intrusion into behavior or into the decoded neural state. For BIND, the analogous manipulation would alter the relation between two patches, $\text{do}(\kappa \mapsto \kappa')$, while preserving each patch's own component-level activity, predicting a selective reduction in conjunction-specific information, $\text{do}(\kappa \mapsto \kappa') \Rightarrow I(J_\kappa; T \mid X, Y) \downarrow$, without a corresponding disruption of either component considered alone. Neither manipulation has yet been reported in the literature this chapter draws on.

This distinction sets up a cleaner transition than a looser reading of the evidence would allow. This chapter has established that cortical space carries behaviorally relevant organization: where a patch of slow-rhythm power sits is not incidental to what a population does next. The next chapter

turns to a different and complementary question, whether the electric fields these patches generate can causally influence subsequent neuronal activity at all, independent of whether that influence, once established, does the specific relational work BIND and REFUSE require. Even granting both chapters' findings together, a further gap remains. Showing that cortical space is organized, and showing separately that fields can act back on spiking, does not yet show that the particular spatial geometry measured here causes the particular operator-level transformations defined in the previous chapter. That composition, between a demonstrated causal channel and a demonstrated operator-level consequence, is exactly what remains missing, and Chapter 8 names it directly as the speculative hinge the rest of the book has to keep in view.

The chapter's central conclusion is accordingly narrower than either an enthusiastic or a dismissive reading would produce. Spatial computing does not supply a decorative map laid over cognition after the fact, since the patches are behaviorally predictive and specifically spatial rather than merely temporally organized. It also does not yet supply a demonstrated field computation, since neither the counterfactual precision REFUSE requires nor the relational information BIND requires has been established by the interventions performed so far. What it supplies, cleanly stated, is a measurable mesoscopic organization: evidence that cortical position itself, not merely which neurons fire or when, functions as a task-relevant variable. That is a substantial finding in its own right, and it is exactly as much as the evidence currently supports.

Chapter 7

Ephaptic Coupling and the Loop Back Into Spiking

Every chapter so far has treated the admissibility field as something that could, in principle, be read off of neural activity: a pattern decodable from local field potentials, correlated with task state, predictive of behavior. None of that evidence, however well established, speaks to a different and prior question. Does the field do anything to the neurons generating it, beyond being an emergent readout of what they are already doing independently? If the field is purely epiphenomenal in this specific sense, a faithful summary statistic of synchronized spiking with no return influence on that spiking, then even a perfectly predictive field measure would be evidence for tier 1 alone, permanently, no matter how precisely it tracked behavior. A causal loop back from field to spiking is a necessary, though not sufficient, condition for anything resembling tier 2.

Ephaptic coupling is the candidate mechanism for that loop. The term refers to a well-established biophysical phenomenon in which the extracellular electric field generated by the coordinated activity of many neurons alters the membrane potential of nearby neurons directly, through volume conduction, rather than through synaptic transmission. Pinotsis and Miller, in a 2023 study [12], present evidence consistent with ephaptic coupling contributing to memory network formation, in which the field generated by an active population appears to help recruit or stabilize the participation of nearby neurons in a shared representation. A related 2026 study from the same authors [13] extends this to variability in neural activity more broadly, presenting evidence that ephaptic influence helps account for trial-to-trial variation in spiking that synaptic input and intrinsic cellular properties alone do not fully explain.

This evidence matters, and it should not be undersold. It converts the field from a pattern that is merely read out of neural activity into a pattern that plausibly acts back upon it, closing a loop that the correlational spatial-computing evidence in the previous chapter could not, on its own, close. If a population's field genuinely influences which nearby neurons participate in a synchronized event, and does so in a way that is not reducible to the synaptic connectivity already driving those neurons, this is a real causal channel through which something like the admissibility field A_t could operate on z_t directly, as the field hypothesis from Chapter 2 requires.

It should also not be oversold, and two distinct cautions apply. The first concerns magnitude. Ephaptic fields generated by ordinary levels of cortical activity are small, typically well under a millivolt at the relevant spatial scale,[1] and the biophysical literature on this question, including work

[1] Anastassiou and colleagues' in vitro and in vivo measurements [1] place typical endogenous extracellular field strengths in the range of a few tenths of a millivolt per millimeter under normal activity, sufficient to produce measurable but modest shifts in spike timing rather than to independently drive spiking in a neuron otherwise near

by Anastassiou and Koch [1] predating the papers this chapter centers on, has long debated whether fields of this size are large enough to meaningfully shift spike timing under realistic physiological conditions, as opposed to the more controlled or synchronized conditions under which the effect is easiest to demonstrate experimentally. The existence of a nonzero ephaptic effect is not in serious dispute. Whether that effect is large enough, under ordinary behaving conditions, to do the amount of selective work this book's later chapters will need it to do is a separate and still open quantitative question, and honesty requires stating it as such rather than letting a demonstrated nonzero effect quietly stand in for a demonstrated sufficient one.

The second caution is more specific to this book's own formal apparatus, and it is the more important of the two. A causal loop from field to spiking, established in general, is not the same claim as a demonstration that this loop implements the particular, content-selective operations Chapter 5 defined. REFUSE requires that a candidate excluding process selectively suppress the admissibility of a specific distinction, while leaving others available; the ephaptic evidence reviewed here shows that fields can influence spiking generally, not that they do so with the selectivity REFUSE's counterfactual condition demands. BIND requires that a joint spatial or phase-based organization carry information about a task-relevant conjunction beyond what its components carry separately; the ephaptic evidence shows a mechanism by which coordinated activity could in principle be maintained or reinforced, not that the specific coordination measured in Chapter 6 carries the synergistic information BIND's criterion requires. Put in the vocabulary the previous chapter used explicitly to flag this exact gap, showing that a causal channel exists between field and spiking, and showing separately that cortical space carries behaviorally predictive organization, does not yet compose into a demonstration that the geometry measured in one study causes the operator-level transformation defined in another. Ephaptic coupling supplies the missing arrow in that composition in principle. It has not yet been shown to supply it for the specific content-selective transformations this book's calculus is built around.

This leaves the field hypothesis in a genuinely intermediate position, and it is worth being precise about exactly what that position is rather than rounding it up or down. The claim that oscillatory fields are purely epiphenomenal, with no causal purchase on subsequent spiking whatsoever, is difficult to sustain against the ephaptic evidence reviewed here. The claim that fields have been shown to perform the specific, content-selective admissibility operations this book formalizes is not yet supported by that same evidence, which establishes a general mechanism rather than a demonstrated instance of the mechanism doing this particular work. Tier 2, recall, was defined as a wave property causally regulating a specific operation, not merely as a wave property causally influencing spiking in some unspecified way. The ephaptic literature makes tier 2 newly plausible as a matter of mechanism. It does not yet establish tier 2 as a matter of demonstrated fact for POP, REFUSE, BIND, or COLLAPSE individually.

The next chapter draws together everything this chapter and the one before it have and have not shown, and names directly the composition that remains missing: a causal mechanism from field to spiking exists in principle, a spatially organized geometry exists and predicts behavior, and yet no experiment reviewed so far has shown that manipulating that geometry, specifically, through that mechanism, specifically, produces the specific operator-level transformations this book has taken pains to define with enough precision that such a demonstration would be recognizable if it occurred. That gap is the speculative hinge the rest of this book has to keep in view, and Chapter 8 states the experiment that would close it.

---

threshold. Strongly synchronized states, such as those seen in some seizure activity, produce substantially larger fields, which is part of why demonstrations of a robust ephaptic effect have often relied on such states or on comparably strong artificial stimulation.

Chapter 8

The Speculative Hinge: Is Interference Itself Computing?

Three chapters have now been spent establishing what a specific, disciplined body of evidence does and does not show, and the pattern across all three has been the same. Spatial patterning of alpha and beta power is real, decodable, and behaviorally predictive, but does not by itself demonstrate the counterfactual exclusion REFUSE requires or the relational information BIND requires. Ephaptic coupling makes it newly plausible that fields act back on the spiking that generates them, but has not been shown to do so with the content-selectivity this book's operators demand. Taken together, the evidence solidly occupies tier 1 for several of the operators, offers a mechanism that could in principle support tier 2, and has not, at any point so far, produced a demonstrated instance of tier 2 for a specific operator, let alone approached tier 3. This chapter states plainly what would have to be shown to move further, and why nothing short of that would count.

The distinction worth holding fixed before stating the experiment is between two claims that are easy to blur in either direction. The weaker claim is that oscillatory fields are informative about, and plausibly causally entangled with, the organization of neural computation: they correlate with task state, they predict behavior, and a real biophysical mechanism exists by which they could influence the spiking they arise from. Nothing in this book disputes that claim, and the preceding three chapters have been devoted to establishing exactly how much of it the current evidence supports. The stronger claim, the one this chapter is named for, is that the geometry of wave interference, specifically the pattern produced when waves of different frequency, phase, and spatial origin overlap and interact, is itself performing a transformation on information, in a sense not reducible to a description of the underlying spiking and recurrent synaptic dynamics. This is analog computation in the strong sense the source literature for this book proposes,[1] and it is the claim on which the book's title, and its distinctiveness against the recurrent-network competitor from Chapter 3, ultimately depends.

The difference between these two claims is not a matter of degree. A system in which fields are informative and causally entangled, but not independently computational, is one in which the field is, in effect, a very useful readout and a partial causal participant in a computation whose essential

[1] "Analog computation" is used elsewhere in the literature, including in some of the source review's own framing, to mean simply computation implemented by continuous rather than discrete physical variables, a much weaker claim that recurrent-network models satisfy just as well as field-based ones. The strong sense at stake in this chapter is narrower: that the specific geometry of wave interference, not merely the continuity of the underlying variables, carries computational content. Readers encountering the term elsewhere should check which of the two senses is intended.

work is still done by spiking and recurrent connectivity, with the field contributing something more like a modulating bias than a distinct computational operation. A system in which interference is genuinely computational is one in which the specific geometry of overlap, the phase relationships and propagation directions among simultaneously active waves, carries information and performs transformations that a complete description of spiking and recurrent state, without that geometry, would fail to capture. Only the second claim licenses saying that the brain uses wave interference as a distinct computational resource, rather than saying that wave measurements are a useful lens onto a computation implemented elsewhere.

The decisive experiment separating these two possibilities has to intervene on geometry specifically, and it has to do so while holding fixed everything a recurrent-network or spiking account would need in order to explain the same outcome without appeal to geometry at all. Concretely, this means perturbing the spatial phase gradient of a traveling wave, its direction and speed of propagation, the completeness of a rotational pattern of the kind Chapter 10 will examine directly in the context of recovery from distraction, or the alignment between two interacting frequency bands, while holding approximately constant the total firing rate of the relevant population, the overall spectral power in each frequency band taken independently of its spatial or phase structure, the sensory input reaching the system, and the difficulty of the task being performed. If behavior, or a decoded operator-level transformation such as POP, REFUSE, BIND, or COLLAPSE as defined in Chapter 5, tracks the manipulated geometry specifically, changing when and only when the geometry changes, while these other quantities are held fixed, the field account gains real causal force and tier 3 becomes newly reachable. If, on the other hand, geometry can be perturbed in this way without any corresponding change in behavior or in the decoded operators, once firing rate, spectral power, sensory input, and task difficulty are properly accounted for, the strong claim fails, and the theory must contract back to the more modest, and still worthwhile, claim that fields are informative measurements and plausible partial causal participants rather than an independent computational layer.

This criterion is deliberately not confined to a single demonstration and is not treated as satisfied once by a favorable result in one paradigm. It recurs at two further points in this book, each time applied to a different empirical setting rather than restated in the abstract. Chapter 11 asks whether the completeness of the rotating wave documented in the distraction-and-recovery literature specifically predicts successful continuation of a train of thought, in a way that firing-rate and recurrent-state accounts of the same recovery cannot equally well explain. Part III applies the criterion to recovery and to the eventual resolution among post-perturbation continuations; continuation failure remains distinct from COLLAPSE, which names a specific resolution among admissible alternatives rather than the loss of an organized regime. Chapter 16 asks whether the specific geometric signature of anesthetic-induced destabilization, rather than merely the fact that some large-scale disruption has occurred, is what tracks loss of behavioral responsiveness across chemically distinct anesthetics, which is this same criterion applied at the scale of a global integrative regime rather than a single operator. Neither of those chapters is permitted to treat a correlation, however striking, as satisfying this bar. Both are required to ask the same geometry-specific, confound-controlled question this chapter has now stated once, precisely, so that it need not be diluted when it is asked again.

It is worth being explicit about what stance this book takes toward its own central hypothesis at this point, roughly the midpoint of the argument. The hypothesis that wave interference performs genuine, irreducible computation has not been established by anything presented so far, and this chapter has not tried to argue that it has. What has been established is a chain of increasingly specific and increasingly plausible claims: a real underdetermination problem that some fast organizing variable must resolve, a formal vocabulary precise enough to say what would count

as that variable doing specific kinds of work, a body of spatial evidence showing that cortical geometry is behaviorally relevant, and a body of biophysical evidence showing that fields are not merely epiphenomenal readouts. None of this proves the strong claim. All of it makes the strong claim worth taking seriously enough to state the experiment that would settle it, and to carry that experiment's absence honestly through the remainder of the argument rather than letting the accumulated weight of suggestive findings substitute for the one kind of finding that would actually decide the question.

Part III

Recovery, Transport, and Holonomy

Chapter 9

What Distraction Interrupts

Everything Part II established concerned organization that is ongoing and largely uninterrupted: task rules held in place, content expressed and bound while a trial proceeds without incident. A different kind of evidence becomes available when that organization is deliberately broken. If cognition depends on a continuously maintained admissibility field rather than on a stored symbolic object that can simply be reopened, then interrupting the field and later observing whether, and how, the interrupted computation resumes is one of the most direct empirical windows available onto what the field is actually doing. This chapter introduces that evidence on its own terms, without yet reaching for the formal machinery its interpretation will eventually require.

The paradigm itself is simple to state. An animal is engaged in a working-memory task, holding some task-relevant content, a remembered item, a rule, a planned response, across a delay period. Partway through that delay, an irrelevant stimulus, a distractor unrelated to the task at hand, is presented.[1] The distractor reliably evokes its own transient pattern of activity, since it is a salient sensory event and the system responds to it as such. The scientific question is what happens next: does the task-relevant content survive this intrusion, and if so, by what process does the system return to whatever it was doing before the interruption occurred.

Batabyal and colleagues, in a 2025 study examining state-space trajectories and traveling waves following distraction [3], give this question a precise experimental answer. Before the distractor appears, the population's neural trajectory can be tracked as it evolves through a state space appropriate to the task, call this pre-distraction trajectory γ^-, ending at some state z^- at the moment the distractor arrives. The distractor then evokes its own activity, displacing the population's state away from wherever the undisturbed trajectory would have carried it. Some time later, a post-distraction state z^+ can be measured, and the animal either continues to perform the task correctly, indicating that the relevant content survived the interruption, or makes an error, indicating that it did not, at least not in a form still usable for the task.

The finding that gives this paradigm its theoretical weight is what happens in the interval between the distractor and the animal's eventual response. Traveling waves are observed following the distractor, and a specific form recurs across trials: a wave pattern that rotates, sweeping in an organized rotational fashion across the relevant cortical territory rather than switching on and off uniformly everywhere at once. The completeness of this rotation, meaning how fully it proceeds through its characteristic cycle before settling, is predictive, on a trial-by-trial basis, of whether the animal's performance following the distraction is correct or erroneous. Trials with a more complete

[1] The distractor is typically chosen to be salient enough to reliably evoke a sensory response and, in some designs, to be superficially similar to the class of items the task uses, so that any observed protection of the target from interference cannot be attributed simply to the distractor being trivially easy to ignore.

rotation are more likely to be followed by successful task performance; trials where the rotation is truncated or fails to develop fully are more likely to be followed by an error.

This finding invites a description that is worth stating carefully before it is examined further, because the natural language for describing it, "the system recovers its train of thought," carries more theoretical commitment than the raw observation supports on its own. What has actually been measured is a statistical relationship between a specific, geometrically characterized feature of post-distraction neural dynamics and a behavioral outcome. Whether this relationship should be read as a wave literally steering the population's state back toward where the pre-distraction trajectory would have led it, as opposed to being a byproduct of some other process, perhaps ordinary recurrent-network relaxation back toward an attractor, that happens to produce a rotational field signature as an incidental consequence, is exactly the kind of question Chapter 3's competitor and Chapter 8's decisive-experiment criterion were built to keep open rather than settle by the choice of descriptive language.

What this chapter can establish, honestly and without yet reaching for that further interpretation, is the shape of the problem the rest of Part III has to solve. A naive notion of recovery, in which z^+ simply equals z^-, the post-distraction state returning exactly to the pre-distraction one, is almost certainly both empirically false and theoretically the wrong thing to ask for. Neural population states are never measured to return to bit-for-bit identical configurations even across ordinary trial repetitions with no distraction at all, given noise, drift, and the sheer dimensionality of the state being measured. A workable notion of recovery cannot require pointwise return. It has to specify some weaker sense in which the post-distraction state counts as a successful continuation of the pre-distraction trajectory, one that captures what the animal's correct subsequent behavior actually requires, without demanding an identity the system was never going to exhibit in the first place. Constructing that weaker notion precisely, and only then asking whether the rotating-wave evidence satisfies it in a way a recurrent-network account of relaxation cannot equally well explain, is the task of the next two chapters.

Chapter 10

Task-Relative Equivalence, Not Literal Return

The previous chapter ended with a negative conclusion: successful recovery from distraction cannot reasonably require the post-distraction state z^+ to equal the pre-distraction state z^-. This chapter supplies the positive replacement. A system has recovered when its post-distraction state preserves the distinctions needed to continue the task, even if its neural coordinates, microscopic configuration, and route through state space differ from those that would have occurred without interruption. Recovery is therefore equivalence relative to a continuation, not identity of physical state.

This replacement is not a concession made merely because exact neural return is difficult to observe. Exact return would often be irrelevant even if it could occur. A working-memory task does not ordinarily require every feature of the pre-distraction population state to be reproduced. It requires preservation of some restricted information: which item was presented, which rule remains operative, which response remains appropriate, or which future distinctions must still be made. Differences outside that task-relevant structure may be large without constituting failure. Conversely, two states may be close under a generic neural distance while differing in the one dimension that determines the correct response. Physical proximity and successful continuation are therefore neither equivalent nor reliably ordered.

Let $\mathcal{T}$ denote a task, understood not merely as an instruction presented to an animal but as a set of future discriminations and actions by which successful continuation is defined. Associate with $\mathcal{T}$ a task map

$$\pi_{\mathcal{T}} : Z \longrightarrow Y_{\mathcal{T}},$$

where Z is the neural state space and $Y_{\mathcal{T}}$ is a task-relevant space of retained contents, decision variables, policies, or predicted outcomes. The map $\pi_{\mathcal{T}}$ discards neural variation that does not matter to the continuation under study and preserves variation that does.

Two neural states $z_1, z_2 \in Z$ are exactly task-equivalent when they induce the same task-relevant state:

$$z_1 \sim_{\mathcal{T}} z_2 \quad \Longleftrightarrow \quad \pi_{\mathcal{T}}(z_1) = \pi_{\mathcal{T}}(z_2).$$

Because equality in $Y_{\mathcal{T}}$ is reflexive, symmetric, and transitive, $\sim_{\mathcal{T}}$ is an equivalence relation. It partitions the neural state space into classes

$$[z]_{\mathcal{T}} = \left\{ z' \in Z : \pi_{\mathcal{T}}(z') = \pi_{\mathcal{T}}(z) \right\}.$$

Each class contains potentially many neural realizations that differ physically while remaining indistinguishable with respect to the task's required continuation.

This definition makes the central claim of the chapter precise:

$$z^+ \sim_{\mathcal{T}} z^-$$

can hold even when

$$z^+ \neq z^-.$$

Successful recovery requires return to the appropriate task-equivalence class, not return to an identical point.

The choice of $\pi_{\mathcal{T}}$ is substantive. It cannot be defined after observing the animal's response so that every correct trial automatically counts as recovered. If the map is fitted using the same behavioral outcome it is then asked to explain, task-equivalence becomes circular. The relevant dimensions must instead be specified independently through the task design, estimated on held-out trials, or recovered from neural activity before distraction and tested after it. A valid map should predict more than the final binary response where the experiment permits. It should preserve the remembered content, the operative rule, the structure of the expected response, or the system's sensitivity to a later probe.

A particularly strong construction uses future discriminability. Let $\mathcal{U}_{\mathcal{T}}$ denote the set of task-relevant probes or perturbations the system may encounter after the candidate recovery point. States z_1 and z_2 are continuation-equivalent when they support the same conditional responses to every relevant future probe:

$$z_1 \sim_{\mathcal{T}} z_2 \quad \Longleftrightarrow \quad p(a \mid z_1, u, \mathcal{T}) = p(a \mid z_2, u, \mathcal{T})$$

for all $u \in \mathcal{U}_{\mathcal{T}}$ and all admissible actions or task outputs a.

This form captures more than agreement on the response that happened to occur. Two states may yield the same immediate choice by coincidence while differing in how they would respond to a later change of cue, a confidence query, a reversal of mapping, or a partial probe of the remembered item. If those future differences matter to the task, the states are not equivalent even if the observed trial ends identically. Continuation is defined by a structured space of possible next demands, not by one realized endpoint alone.

In practice, exact equality of conditional response distributions will rarely be estimable. An approximate relation is therefore needed. Let $d_{\mathcal{T}}$ be a metric or divergence on task-relevant states. We may write

$$z_1 \approx_{\mathcal{T},\delta} z_2 \quad \Longleftrightarrow \quad d_{\mathcal{T}}(\pi_{\mathcal{T}}(z_1), \pi_{\mathcal{T}}(z_2)) \leq \delta,$$

where δ is fixed by the resolution required for successful continuation.

Unlike the exact relation $\sim_{\mathcal{T}}$, the tolerance relation $\approx_{\mathcal{T},\delta}$ need not be transitive.[1] A state may lie within δ of a second state and the second within δ of a third while the first and third lie farther apart than δ. The distinction matters later, when recovery is discussed as transport through a sequence of local states. Approximate local preservation does not automatically guarantee global preservation around a long trajectory. Accumulated deviation can remain invisible at each small step and still become task-relevant by the end.

[1] This is the familiar failure of transitivity in any tolerance relation defined by a threshold on a metric, sometimes discussed under the heading of the sorites-like structure of similarity judgments. It is not a defect specific to this construction; it is a generic feature of any δ-ball criterion, and the reason exact equivalence classes, despite being the less realistic idealization, remain useful as the reference point against which tolerance-based approximations are checked.

State Recovery and Trajectory Recovery

Returning to a task-equivalence class at one moment is not always enough. A post-distraction state may decode as task-correct briefly and yet enter dynamics that immediately carry it away from successful performance. Conversely, a state initially displaced from the pre-distraction class may lie on a trajectory that reliably restores the relevant information before it is needed. Recovery should therefore be evaluated both as a property of states and as a property of trajectories.

Let

$$\widehat{\gamma}^{0}_{[t_d,t_1]}$$

denote the counterfactual trajectory predicted to have occurred from distractor onset t_d to a later task-relevant time t_1 had no distraction been presented. Let

$$\gamma^{d}_{[t_d,t_1]}$$

denote the observed trajectory after distraction. The superscripts 0 and d mark the undistracted counterfactual and distracted observation respectively.

A strict trajectory comparison in the full neural state space would recreate the original problem. The distracted and undistracted paths may differ substantially while preserving the same task-relevant continuation. The comparison must therefore be made after projection:

$$\pi_{\mathcal{T}} \circ \gamma^{d} \qquad \text{and} \qquad \pi_{\mathcal{T}} \circ \widehat{\gamma}^{0}.$$

A task-relative trajectory distance can be defined as

$$D_{\mathcal{T}}\Big(\gamma^{d}, \widehat{\gamma}^{0}\Big) = \inf_{\tau \in \mathfrak{T}} \int_{t_d}^{t_1} w(t)\, d_{\mathcal{T}}\Big(\pi_{\mathcal{T}}(z_t^{d}), \pi_{\mathcal{T}}(\widehat{z}^{0}_{\tau(t)})\Big)\, dt.$$

Here $w(t)$ weights moments by their importance to the task, and $\mathfrak{T}$ is a restricted family of admissible temporal alignments. The infimum permits modest differences in recovery speed without declaring failure merely because two otherwise equivalent trajectories are traversed at slightly different rates. The family $\mathfrak{T}$ must be constrained in advance; allowing arbitrary time warping would make almost any sufficiently flexible path appear recovered.

A normalized recovery score can then be written

$$R_{\mathcal{T}} = 1 - \min\left\{1,\ \frac{D_{\mathcal{T}}\big(\gamma^{d}, \widehat{\gamma}^{0}\big)}{\Lambda_{\mathcal{T}}}\right\},$$

where $\Lambda_{\mathcal{T}} > 0$ is a task-appropriate failure scale. Thus

$$0 \le R_{\mathcal{T}} \le 1.$$

Values near one indicate close task-relative agreement with the predicted uninterrupted continuation; values near zero indicate deviation at or beyond the scale associated with functional failure.

The normalization $\Lambda_{\mathcal{T}}$ should not be an arbitrary maximum distance chosen for convenience. It may be estimated from the separation between successful and failed trajectories, from the distance at which future task performance falls to chance, or from a preregistered behavioral tolerance. Different choices answer different questions. A scale based on correct-versus-error separation measures behavioral recoverability. A scale based on undistracted trial variability measures return to the normal task regime. These should not be silently substituted for one another.

The score is graded because recovery itself may be graded. A remembered item can return with reduced precision. A rule can be restored while its associated confidence remains damaged. A trajectory can regain the correct response basin but lose information that would have supported an alternative probe. Binary success at the end of the trial compresses these differences. $R_{\mathcal{T}}$ is intended to preserve them.

The Counterfactual Problem

The uninterrupted trajectory

$$\widehat{\gamma}^{\,0}$$

is not directly observable on a distracted trial. The same system cannot both receive and not receive the distractor in the same trial. Any recovery score that compares the observed path with an uninterrupted path therefore depends upon a counterfactual estimate, and the reliability of the score cannot exceed the reliability of that estimate.

The simplest estimator averages trajectories from matched undistracted trials. This is inadequate when neural dynamics vary substantially from trial to trial, because the mean path may be a trajectory no individual trial actually follows. A stronger estimator conditions the predicted continuation on the pre-distraction history:

$$\widehat{\gamma}^{0}_{[t_d,t_1]} = \mathbb{E}\Big[\gamma^{0}_{[t_d,t_1]} \mid z_{\le t_d}, u_{\le t_d}, \mathcal{T}\Big].$$

The prediction uses the observed trajectory up to distraction onset to estimate where that particular trial would otherwise have gone.

Even this conditional expectation can obscure multimodality. If the undisturbed system could have followed several distinct but successful continuations, their average may pass through a region corresponding to none of them. The counterfactual target should then be a distribution over admissible uninterrupted trajectories:

$$p\Big(\gamma^{0}_{[t_d,t_1]} \mid z_{\le t_d}, u_{\le t_d}, \mathcal{T}\Big),$$

and recovery should be measured by the distance from the observed post-distraction trajectory to that distribution or to its task-equivalent support, rather than to a single mean path.

One possible distributional score is

$$R^{\mathrm{dist}}_{\mathcal{T}} = \Pr\Big(D_{\mathcal{T}}\Big(\gamma^{\,d}, \Gamma^{0}\Big) \le \Lambda_{\mathcal{T}} \;\Big|\; z_{\le t_d}, \mathcal{T}\Big),$$

where Γ^0 is a random uninterrupted continuation drawn from the estimated counterfactual distribution. This asks whether the distracted trajectory falls within the range of task-relevant continuations the system plausibly could have followed without distraction.

The counterfactual model must be trained without using post-distraction outcome labels in a way that leaks success or failure into the target. It should also be validated on held-out undistracted trials by pretending that a distraction occurred and testing whether the model correctly predicts the continuation. Without that validation, a small recovery distance may reflect an overly broad model rather than genuine return, while a large distance may reflect poor counterfactual prediction rather than failure of the neural system.

Recovery Is Not Necessarily Reversal

The language of return can suggest that the system must undo the distraction by tracing its path backward. Nothing in task-relative equivalence requires this. Let the distractor carry the system from z^- to a displaced state z^d. Successful continuation may follow a new route:

$$z^- \longrightarrow z^d \longrightarrow z^+,$$

where $z^+ \sim_{\mathcal{T}} \widehat{z}^0$ even though the segment from z^d to z^+ is not the reverse of the segment from z^- to z^d.

This distinction separates restoration from rewind. A dissipative dynamical system rarely retraces its microscopic history. It may instead converge upon the same task-relevant manifold from another direction. A rotating traveling wave may participate in such convergence without storing or replaying the pre-distraction state point by point. Its possible role is to transport the system toward an equivalent orientation from which the task can continue.

Nor must successful recovery restore every operation interrupted by the distractor. Suppose a bound item-location representation is partially dissolved while the item identity remains available. If the eventual probe asks only for item identity, the animal may respond correctly without having restored the original binding. Relative to that probe, recovery occurred; relative to the richer task state that existed before distraction, it did not. We may write

$$z^+ \sim_{\mathcal{T}_1} z^-, \qquad z^+ \not\sim_{\mathcal{T}_2} z^-,$$

where $\mathcal{T}_1$ tests item identity and $\mathcal{T}_2$ tests the item-location conjunction.

Recovery claims are consequently indexed by what the experiment tests. A correct response licenses only the claim that enough structure survived or was reconstructed for that response. It does not establish preservation of the entire preceding thought. Richer probe sets make the equivalence class narrower and the recovery claim stronger.

Persistence as Preservation of Future Distinctions

Task-relative equivalence supplies a first formal meaning for persistence. A distinction persists through distraction when the transformation induced by that distraction leaves intact the future discriminations that define its continuation. Let

$$D_{[t_d,t_r]} : Z \longrightarrow Z$$

denote the net transformation from distraction onset t_d to a candidate recovery time t_r. Persistence of z^- relative to task $\mathcal{T}$ requires

$$D_{[t_d,t_r]}(z^-) \sim_{\mathcal{T}} \widehat{z}^0_{t_r}.$$

Equivalently,

$$\pi_{\mathcal{T}}\big(D_{[t_d,t_r]}(z^-)\big) = \pi_{\mathcal{T}}\big(\widehat{z}^0_{t_r}\big)\,.$$

The transformation need not preserve the original state. It need preserve only the quotient defined by $\pi_{\mathcal{T}}$. In algebraic language, successful recovery allows substantial motion within an equivalence class while forbidding motion that crosses a task-relevant class boundary. This is the first point at which the book's broader claim about continuation becomes exact: persistence is not invariance of material configuration but invariance of the distinctions upon which an admissible future depends.

The formulation also permits partial persistence. Let

$$\pi_{\mathcal{T}} = (\pi_1, \pi_2, \ldots, \pi_k)$$

decompose the task map into relevant discriminations. The system may preserve some components and lose others:

$$\pi_i(z^+) = \pi_i(\widehat{z}^0) \quad \text{for } i \in K,$$

while

$$\pi_j(z^+) \neq \pi_j(\widehat{z}^0) \quad \text{for } j \notin K.$$

A weighted recovery score can reflect their unequal importance:

$$R_{\mathcal{T}}^{\text{disc}} = \sum_{i=1}^{k} \omega_i r_i, \qquad \omega_i \geq 0, \qquad \sum_{i=1}^{k} \omega_i = 1,$$

where r_i measures preservation of the ith distinction. This provides a more interpretable decomposition than a single global neural distance whenever the task design exposes several separable contents or relations.

What the Rotating Wave Evidence Establishes

The Batabyal result can now be assigned a precise place within this framework. A more complete post-distraction rotation predicts a greater probability of correct task performance. If rotational completeness is denoted by C_{rot}, the reported relationship has the form

$$\Pr\Big(\widehat{T} = T \mid C_{\text{rot}} = c_2\Big) > \Pr\Big(\widehat{T} = T \mid C_{\text{rot}} = c_1\Big)$$

for appropriately ordered values $c_2 > c_1$.

This supports the claim that rotational geometry is informative about successful continuation. If task-relevant neural decoding also returns toward the undistracted regime as rotation completes, then the evidence supports an association between C_{rot} and a neural recovery score such as

$$I(C_{\text{rot}}; R_{\mathcal{T}} \mid C) > 0,$$

where C includes measured confounders.

That remains a tier-1 result unless the geometry is independently manipulated or a comparably strong causal identification strategy is available. A complete rotation may be the mechanism that restores the task-equivalence class. It may instead be generated by recurrent relaxation that independently restores both neural state and behavior. It may be a marker of intact arousal or network stability that makes recovery possible without itself directing the recovery. Task-relative equivalence tells us what successful recovery means; it does not, by itself, tell us what causes it.

The decisive experiment from Chapter 8 therefore remains in force. Rotational geometry would clear tier 2 for recovery if manipulating its completeness or direction altered $R_{\mathcal{T}}$ while preserving the relevant component activity and background state:

$$p(R_{\mathcal{T}} \mid \text{do}(C_{\text{rot}} = c_2), C) \neq p(R_{\mathcal{T}} \mid \text{do}(C_{\text{rot}} = c_1), C)\,.$$

It would approach tier 3 only if geometry retained explanatory and causal force beyond the strongest adequately measured recurrent-state model:

$$I\big(R_{\mathcal{T}}; C_{\text{rot}} \mid R_t^{\text{recurrent}}, u_t, C\big) > 0,$$

combined with a geometry-selective intervention whose effect could not be explained by changes in firing rate, spectral power, or recurrent mediation.

From Equivalence to Transport

This chapter has replaced literal return with a task-relative equivalence relation. The replacement allows recovery to be graded, makes its counterfactual target explicit, distinguishes state recovery from trajectory recovery, and defines persistence through the future distinctions a system remains able to make. It also leaves an important question unanswered.

The relation

$$z^+ \sim_{\mathcal{T}} \hat{z}^0$$

compares an endpoint with its task-relevant target. It does not describe how the system transports the relevant distinctions through the intervening perturbation, nor how local changes along a path can accumulate into either preserved or altered orientation. Two trajectories may begin and end in the same task-equivalence classes while differing profoundly in how they carry representational relations between them. Conversely, every short segment of a path may look approximately preserving while the completed circuit returns with a residual transformation.

Those are questions about transport rather than endpoint equality. They require a language for identifying which transformations count as changes of neural realization and which count as changes in the distinctions being preserved. The next chapter introduces that language through distinction holonomy and the gauge of persistence. It does so only now, after the empirical recovery problem and the equivalence relation it requires have both been specified independently. Holonomy will not be used as a metaphor for rotation. It will be asked to earn a narrower role: measuring whether transport around an interruption returns task-relevant distinctions with their operative relations intact.

Chapter 11

Distinction Holonomy and the Gauge of Persistence

The previous chapter defined recovery without requiring a return to an identical neural state. Two states count as equivalent when they preserve the distinctions needed for the same task-relative continuation:

$$z_1 \sim_{\mathcal{T}} z_2 \quad \Longleftrightarrow \quad \pi_{\mathcal{T}}(z_1) = \pi_{\mathcal{T}}(z_2).$$

That relation identifies what may vary without constituting cognitive loss. It does not yet describe how a task-relevant organization is carried through time, how representational coordinates change along a trajectory, or how a sequence of locally acceptable changes can accumulate into a globally consequential residual. Those are transport questions.

The language of holonomy is appropriate only if four objects can be specified independently: a state space, a family of fibers containing alternative realizations, a transformation group relating those realizations, and a connection specifying how representations are compared across changing contexts. Without those objects, writing a path-ordered exponential would add mathematical appearance without mathematical content. This chapter constructs them in that order.

The construction is deliberately conditional. It does not claim that cortical activity has already been shown to instantiate a unique biological gauge theory. It asks whether the experimentally motivated equivalence relation from Chapter 10 admits a useful gauge description, what measurements such a description would require, and what additional empirical work would distinguish genuine transport structure from an elaborate redescription of ordinary state-space dynamics.

The Neural State Space and the Task Base

Let Z denote the neural state space at the scale of observation relevant to the experiment. A point

$$z \in Z$$

may contain population firing rates, latent factors inferred from those rates, local field-potential variables, phase relations, or some preregistered combination of these. The choice of Z is not neutral. A state space constructed only from firing rates cannot test whether phase geometry contributes additional structure, while one constructed only from field measurements cannot control adequately for recurrent population state. The space must be rich enough to represent the competing accounts the analysis is intended to distinguish.

Let $M_{\mathcal{T}}$ denote the base space of task-relative conditions. A point

$$m \in M_{\mathcal{T}}$$

specifies the task circumstances under which a representation must remain usable: the current trial epoch, operative rule, relevant sensory context, expected class of future probes, and other variables needed to define successful continuation. The base is not the cortical surface. Cortical location may enter the neural realization, and traveling waves move across that physical surface, but the base space used here records the task-relative conditions across which representational organization must be transported.

This distinction prevents two geometries from being conflated. There is a physical geometry of cortical position, denoted schematically by X_{cortex}, and there is a task-context geometry $M_{\mathcal{T}}$. A continuation field may depend upon both:

$$\mathcal{A}^{\text{wave}} : X_{\text{cortex}} \times M_{\mathcal{T}} \times \mathbb{R} \longrightarrow \mathcal{V},$$

where $\mathcal{V}$ contains measured field variables such as amplitude, phase, frequency, or propagation direction. The transport construction asks how changes in these physical variables correspond to preservation or alteration of task-relative distinctions as the system moves through $M_{\mathcal{T}}$.

The task map from Chapter 10 can now be indexed by context:

$$\pi_{\mathcal{T},m} : Z_m \longrightarrow Y_{\mathcal{T},m}.$$

Here Z_m is the set of neural states attainable under context m, and $Y_{\mathcal{T},m}$ contains the task-relevant contents, relations, or policies that must be preserved there. Context indexing matters because successful realization need not have the same neural coordinates at encoding, maintenance, recovery, and response. What counts as preservation is not numerical constancy of a decoder output across all epochs, but lawful transport of the represented distinction from the form appropriate at one epoch to the form appropriate at another.

Fibers of Neural Realization

For a task-relevant value $y \in Y_{\mathcal{T},m}$, define the realization fiber

$$F_{m,y} = \{z \in Z_m : \pi_{\mathcal{T},m}(z) = y\}.$$

The fiber contains neural states that differ physically but realize the same task-relevant distinction under context m. If y represents the remembered conjunction "red on the left," then every state in $F_{m,y}$ must preserve whatever future discriminations the experiment uses to distinguish that conjunction from alternatives, even though individual neurons, firing rates, phases, or latent coordinates may differ across states.

The total realization space may be written

$$\mathcal{E}_{\mathcal{T}} = \bigsqcup_{m \in M_{\mathcal{T}}} \bigsqcup_{y \in Y_{\mathcal{T},m}} F_{m,y},$$

with projection

$$p : \mathcal{E}_{\mathcal{T}} \longrightarrow M_{\mathcal{T}}.$$

This notation resembles a fiber bundle, but the resemblance should not be mistaken for a result. A genuine smooth fiber bundle requires local regularity: neighboring contexts must have fibers of compatible structure, and local trivializations must exist. Neural data may violate these assumptions

through bifurcations, abrupt representational remapping, changes in effective dimension, or task epochs with qualitatively different codes. Where such regularity fails, the appropriate object may be a stratified space, a family of local bundles, or merely a collection of context-indexed realization sets.

The bundle language is therefore earned only locally and empirically. One must show that representations can be tracked continuously across a region of task space and that changes of neural coordinates between overlapping regions are sufficiently lawful to be modeled as transformations rather than arbitrary remappings.

The fibers make precise the gauge intuition introduced in Chapter 10. Neural realization contains degrees of freedom that may change while task-relative content remains invariant. Movement within a fiber changes the realization but not the represented distinction:

$$z \mapsto z', \qquad z, z' \in F_{m,y}.$$

Such movement is vertical with respect to the projection. Movement across task contexts,

$$F_{m_0,y_0} \longrightarrow F_{m_1,y_1},$$

is not automatically content-preserving. It requires a rule specifying which realization at the later context counts as the continuation of a realization at the earlier one. That rule is the connection.

Representational Frames and the Transformation Group

The task map alone is not enough to define holonomy. If a fiber is defined strictly as the set of states with identical task output, then every transformation within it preserves that output by construction. Holonomy acting only inside such a fiber could never represent continuation failure. A richer object is required: not merely the decoded content, but the representational frame in which content and its possible future relations are organized.

Let V_m denote a local representational space over context m. A frame

$$e_m = (e_m^1, \ldots, e_m^k)$$

specifies a basis or coordinate system for task-relevant distinctions within that space. The same task content can be represented in different frames. If $v_m \in V_m$ gives the coordinates of a content in frame e_m, a change of frame by

$$g \in G$$

acts as

$$e_m \mapsto e_m g, \qquad v_m \mapsto g^{-1} v_m,$$

leaving the represented object invariant when both transformations are made consistently.

The group G is the candidate transformation group of representational reparameterizations. It should not be chosen merely because a familiar Lie group makes the mathematics convenient. Its form depends upon the empirical representation. If only rotations of a latent neural subspace leave decoding invariant, G may be an orthogonal group. If arbitrary invertible linear changes of coordinates preserve the relevant structure, G may be a subgroup of $\mathrm{GL}(k)$. If phase offsets supply the relevant redundancy, a circular group such as $U(1)$ may appear locally. Nonlinear representational manifolds may require a more general group or groupoid of admissible transformations.

The full group G contains transformations that can relate local representational frames. Not every element of G preserves the distinctions required by task $\mathcal{T}$. Define the task-preserving subgroup

$$K_{\mathcal{T}} = \{g \in G : d_{\mathcal{T}}(gv) = d_{\mathcal{T}}(v) \text{ for all relevant } v\},$$

where

$$d_{\mathcal{T}} : V_m \longrightarrow Y_{\mathcal{T},m}$$

is the task-relative decoding or future-discrimination map.

Elements of $K_{\mathcal{T}}$ alter representational coordinates without changing the task-relevant continuation. Elements of $G \setminus K_{\mathcal{T}}$ change at least one distinction the task requires. This subgroup resolves a problem that would otherwise make the gauge formulation empty. Successful persistence does not require the accumulated transformation to be the identity element of G. It requires only that the transformation lie in the stabilizer of the relevant distinctions:

$$g_{\text{net}} \in K_{\mathcal{T}}.$$

A nonidentity transformation can therefore be cognitively harmless, while a small transformation outside $K_{\mathcal{T}}$ can constitute failure if it crosses a task-relevant boundary.

The subgroup is task-relative. A transformation may preserve item identity while exchanging item-location bindings:

$$g \in K_{\mathcal{T}_1}, \qquad g \notin K_{\mathcal{T}_2}.$$

This is the group-theoretic counterpart of the result from Chapter 10:

$$z^+ \sim_{\mathcal{T}_1} z^-, \qquad z^+ \not\sim_{\mathcal{T}_2} z^-.$$

The same neural transport can be harmless relative to a coarse probe and destructive relative to a more discriminating one.

The Connection as a Rule of Comparison

States recorded at different task epochs need not use the same neural coordinates. A decoder trained during encoding may fail during maintenance even when the memory remains behaviorally intact, because the representation has transformed rather than disappeared. To decide whether such a transformation preserves continuation, one needs a rule for comparing representational frames at neighboring points of the base.

A connection supplies that rule. Denote the task-relative connection by

$$\omega^{(\mathcal{T})}.$$

The symbol ω is used rather than A to avoid confusing the geometric connection with the admissibility set A_t introduced in Chapter 5 or the wave contribution $\mathcal{A}_t^{\text{wave}}$ introduced in Chapter 2.

Locally, if the task context follows a path

$$\Gamma : s \longmapsto m(s)$$

through $M_{\mathcal{T}}$, the connection determines how a representational vector is transported:

$$\frac{Dv}{Ds} = \frac{dv}{ds} + \omega_\mu^{(\mathcal{T})}(m(s))\frac{dm^\mu}{ds}\, v.$$

Parallel transport is defined by

$$\frac{Dv}{Ds} = 0.$$

The equation does not mean that the neural state remains constant. It means that changes in its coordinates are exactly those attributed to lawful change of representational frame along the path.

For a short displacement dm^μ, the transported vector changes approximately as

$$v(m + dm) \approx \left[I - \omega_\mu^{(\mathcal{T})}(m)\, dm^\mu \right] v(m).$$

The connection therefore separates two kinds of observed change: change expected because the representational frame itself is evolving, and residual change that alters the transported distinction relative to that evolving frame.

The connection cannot be inferred merely by choosing whichever transport makes successful trials look invariant. That would recreate the circularity Chapter 10 excluded for the task map. It must be estimated independently, for example from undistracted trials, cross-validated transformations between task epochs, perturbations known not to alter the remembered content, or a generative dynamical model fitted without access to the recovery outcome later being tested.

A viable empirical connection should satisfy at least three requirements. It should transport held-out undistracted representations across task epochs with low task-relative error. It should preserve known distinctions while allowing empirically observed representational drift. It should make predictions about distracted or perturbed trials that are not built into its fitting procedure. If no such transport rule generalizes beyond the data used to define it, the connection is an analytical convenience rather than evidence for persistent representational geometry.

Transport Along an Interruption

Given a connection, transport along a path Γ from context m_0 to context m_1 is represented by

$$U_\Gamma^{(\mathcal{T})} = \mathcal{P} \exp\left(- \int_\Gamma \omega^{(\mathcal{T})} \right).$$

The path-ordering symbol $\mathcal{P}$ is required when infinitesimal transformations at different points do not commute. In that case, the order in which the system encounters changes matters:

$$\omega(s_1)\omega(s_2) \neq \omega(s_2)\omega(s_1).$$

This is a natural possibility for cognition. A distractor followed by a rule cue need not have the same consequence as the same rule cue followed by the distractor, even if the component events are otherwise identical.

For the distraction paradigm, let Γ_d denote the path through task and neural context induced by the distractor and subsequent recovery interval. Then

$$v^+ = U_{\Gamma_d}^{(\mathcal{T})} v^-$$

gives the transported representation predicted by the connection. Successful continuation does not require

$$v^+ = v^-$$

in raw coordinates. It requires that the net transport preserve the task-relative decoding:

$$d_{\mathcal{T}}(v^+) = d_{\mathcal{T}}(\widehat{v}^{\,0}),$$

where $\widehat{v}^{\,0}$ is the counterfactual uninterrupted representation at the same task-relative endpoint.

This is still an open-path comparison. The initial and final contexts differ, and an open-path transport operator is not, by itself, a gauge-invariant object. Under a change of local frame,

$$U_\Gamma^{(\mathcal{T})} \mapsto g(m_1)^{-1} U_\Gamma^{(\mathcal{T})} g(m_0).$$

Its matrix entries depend upon how frames are chosen at the two endpoints. A claim about "the transformation accumulated during recovery" is therefore incomplete unless the endpoint frames are identified by an independently specified comparison.

The counterfactual trajectory from Chapter 10 provides one such identification. Let Γ_0 denote the predicted uninterrupted path from the same initial context to the same task-relative endpoint. The relative transport

$$\Delta U_{\mathcal{T}} = \left(U_{\Gamma_0}^{(\mathcal{T})}\right)^{-1} U_{\Gamma_d}^{(\mathcal{T})}$$

compares distracted transport with the transport expected without distraction. Because the two paths share their endpoints, the relative operator transforms by conjugation at the initial frame:

$$\Delta U_{\mathcal{T}} \mapsto g(m_0)^{-1} \Delta U_{\mathcal{T}} g(m_0).$$

Its conjugacy class, eigenvalue structure where applicable, and action on task-relevant subspaces can therefore carry frame-independent content.

Equivalently, concatenate the distracted path Γ_d with the reverse of the uninterrupted path:

$$C = \Gamma_d \circ \Gamma_0^{-1}.$$

This produces a closed comparison loop in task-context space. Holonomy now becomes appropriate.

Task-Relative Holonomy

The holonomy around the comparison loop C is

$$H_C^{(\mathcal{T})} = \mathcal{P} \exp\left(-\oint_C \omega^{(\mathcal{T})}\right).$$

It measures the net representational transformation accumulated by following the distracted continuation to the endpoint and returning through the counterfactual uninterrupted identification.

The central criterion is not

$$H_C^{(\mathcal{T})} = I.$$

Exact identity would require the representational frame to return without any residual transformation whatsoever, which is the geometric version of the pointwise-return standard rejected in Chapter 10. Successful persistence requires instead that the holonomy act trivially on the distinctions needed for task $\mathcal{T}$:

$$H_C^{(\mathcal{T})} \in K_{\mathcal{T}}.$$

Equivalently,

$$d_{\mathcal{T}}\left(H_C^{(\mathcal{T})} v\right) = d_{\mathcal{T}}(v)$$

for every task-relevant representational vector v.

Continuation failure occurs when the residual transformation leaves the task-preserving subgroup:

$$H_C^{(\mathcal{T})} \notin K_{\mathcal{T}}.$$

In that case, transport around the interruption changes at least one future discrimination required by the task. The failure need not be large in the metric of the full neural state space. A small residual rotation across a narrow decision boundary may matter more than a large transformation confined to task-irrelevant dimensions.

A graded holonomy defect can be defined by the distance from the task-preserving subgroup:

$$\delta_{\mathrm{hol}}^{(\mathcal{T})} = \inf_{k \in K_{\mathcal{T}}} d_G\Big(H_C^{(\mathcal{T})}, k\Big),$$

where d_G is a metric or divergence appropriate to the transformation group. Then

$$\delta_{\mathrm{hol}}^{(\mathcal{T})} = 0$$

indicates exact task-relative preservation, while increasing values indicate a growing residual outside the set of transformations that preserve the required distinctions.

A content-sensitive version measures the action directly:

$$\delta_{\mathrm{act}}^{(\mathcal{T})}(v) = d_{Y_{\mathcal{T}}}\Big(d_{\mathcal{T}}(H_C^{(\mathcal{T})}v), d_{\mathcal{T}}(v)\Big).$$

This avoids treating every group direction as equally important. Two holonomies may be equally distant from the identity in G while having very different consequences for the represented content.

The relation to the recovery score from Chapter 10 is then

$$R_{\mathcal{T}} \approx \Psi\Big(\delta_{\mathrm{hol}}^{(\mathcal{T})}, \delta_{\mathrm{state}}^{(\mathcal{T})}, C\Big),$$

where $\delta_{\mathrm{state}}^{(\mathcal{T})}$ measures task-relative endpoint error and C contains other relevant conditions. Holonomy and endpoint recovery are related but not identical. The endpoint score asks whether the system arrived in an acceptable task-relative state. The holonomy defect asks what transformation accumulated in transporting the representation there.

Curvature and Path Dependence

A connection becomes especially informative when different routes between the same task contexts yield different transported representations. The local measure of this path dependence is the curvature

$$\Omega^{(\mathcal{T})} = d\omega^{(\mathcal{T})} + \omega^{(\mathcal{T})} \wedge \omega^{(\mathcal{T})}.$$

If

$$\Omega^{(\mathcal{T})} = 0$$

throughout a simply connected region, transport there is locally path-independent up to gauge, and sufficiently small contractible loops have trivial holonomy. If

$$\Omega^{(\mathcal{T})} \neq 0,$$

the order and route through contextual changes can produce a residual transformation even when the system returns to the same nominal task condition.

This gives distinction holonomy empirical content beyond the observation that neural states drift. Consider two sequences containing the same component events:

$$\Gamma_{12} : m_0 \to m_1 \to m_2,$$

and

$$\Gamma_{21} : m_0 \to m_1' \to m_2.$$

One might correspond to a distractor followed by a rule reminder, the other to the rule reminder followed by the distractor. If transport is path-dependent,

$$U_{\Gamma_{12}}^{(\mathcal{T})} \neq U_{\Gamma_{21}}^{(\mathcal{T})},$$

and the difference may affect task-relative decoding even though the initial context, final context, and component events are matched.

A genuine test of curvature would therefore compare ordered perturbation sequences rather than merely compare activity before and after one interruption. The relevant prediction is not simply that different event orders yield different neural states; recurrent dynamical systems already predict such order effects. The stronger prediction is that the differences are captured by a transport law whose locally estimated connection predicts the residual around novel loops:

$$H_C^{\text{predicted}} \approx H_C^{\text{observed}}$$

on held-out paths. Without successful out-of-sample loop prediction, curvature remains a suggestive geometrical description rather than an independently supported property.

The Rotating Wave and the Transport Law

The rotating traveling wave observed after distraction is an intuitively attractive candidate for a physical transport process. A spatial rotation has direction, phase progression, extent, and completeness, all of which could contribute to the mapping between representational frames before and after interruption. The temptation is therefore to identify the observed physical rotation directly with holonomy.

That identification would be premature. A rotating pattern in cortical coordinates is not automatically a connection on a representational bundle, and one completed physical rotation is not automatically a closed loop in task-context space. The physical wave and the representational transport occupy related but distinct descriptive domains:

$$G_t^{\text{wave}} \quad \text{and} \quad \omega^{(\mathcal{T})}.$$

The empirical question is whether the former helps determine the latter:

$$\omega^{(\mathcal{T})} = \Phi\big(G_t^{\text{wave}}, R_t^{\text{recurrent}}, u_t, m_t\big)\,,$$

where $R_t^{\text{recurrent}}$ summarizes the recurrent population state.

The rotating-wave hypothesis gains content if wave phase and propagation geometry predict the transport between task-relative representational frames beyond what is predicted by recurrent state alone:

$$I\Big(U_\Gamma^{(\mathcal{T})}; G_t^{\text{wave}} \mid R_t^{\text{recurrent}}, u_t, m_t\Big) > 0.$$

It gains causal force if manipulating rotational geometry changes the holonomy defect:

$$p\Big(\delta_{\text{hol}}^{(\mathcal{T})} \mid \text{do}(C_{\text{rot}} = c_2), C\Big) \neq p\Big(\delta_{\text{hol}}^{(\mathcal{T})} \mid \text{do}(C_{\text{rot}} = c_1), C\Big)\,.$$

It approaches the third evidential tier only if this intervention has task-specific consequences that cannot be accounted for by changes in spectral power, average firing, arousal, or recurrent mediation.

Rotational completeness can then be interpreted as a possible control parameter for transport, not as holonomy itself. A complete physical rotation may implement transport whose residual lies within $K_{\mathcal{T}}$:

$$C_{\text{rot}} \approx 1 \quad \Rightarrow \quad H_C^{(\mathcal{T})} \in K_{\mathcal{T}}.$$

A truncated rotation may leave a residual outside that subgroup:

$$C_{\text{rot}} < c_{\text{crit}} \quad \Rightarrow \quad H_C^{(\mathcal{T})} \notin K_{\mathcal{T}}.$$

These are testable hypotheses, not conclusions already licensed by the reported correlation between rotational completeness and behavioral success.

The direction of implication must also remain open. Successful recurrent recovery may generate a complete rotating wave as its field signature:

$$\text{recurrent recovery} \longrightarrow C_{\text{rot}} \longrightarrow \text{measured field pattern},$$

rather than the wave producing recovery:

$$C_{\text{rot}} \longrightarrow H_C^{(\mathcal{T})} \in K_{\mathcal{T}} \longrightarrow \text{successful continuation}.$$

Only geometry-selective intervention or an equivalently strong causal design can discriminate these possibilities.

Gauge Freedom and Empirical Invariance

A useful gauge description must distinguish coordinate-dependent quantities from claims that should survive changes of representational basis. Under a local change of frame

$$g : M_{\mathcal{T}} \to G,$$

the connection transforms as

$$\omega^{(\mathcal{T})} \mapsto g^{-1}\omega^{(\mathcal{T})}g + g^{-1}dg.$$

The holonomy around a loop based at m_0 transforms by conjugation:

$$H_C^{(\mathcal{T})} \mapsto g(m_0)^{-1} H_C^{(\mathcal{T})} g(m_0).$$

Individual matrix entries of $H_C^{(\mathcal{T})}$ are therefore basis-dependent. Its conjugacy class, traces in an appropriate representation, eigenvalue structure where defined, and action on independently specified task-relevant subspaces are better candidates for invariant empirical claims.

Most importantly, task success must not depend upon the analyst's arbitrary choice of latent-space basis. If two equally valid dimensionality-reduction methods yield coordinate systems related by an admissible transformation, the conclusion about whether transport preserved the task-relative distinction should agree:

$$H_C^{(\mathcal{T})} \in K_{\mathcal{T}}$$

must be invariant under simultaneous transformation of the holonomy, decoder, and task-preserving subgroup.

This provides a practical test of whether gauge language is doing real work. If the measured recovery effect disappears under an innocuous change of representational coordinates, it is unlikely to identify a physically or cognitively meaningful transport property. If a task-relative residual

persists across admissible coordinate choices and predicts future discrimination on held-out trials, the case for a genuine invariant becomes stronger.

Gauge freedom here does not mean that every neural description is equally valid. The admissible transformations are constrained by preserved observables and by the independently defined task map. A transformation that destroys known temporal, anatomical, causal, or decoding structure is not a harmless change of gauge merely because it can be written mathematically. The empirical content lies in determining which transformations genuinely preserve the system's operative distinctions.

Holonomy, REFUSE, BIND, and COLLAPSE

The transport formalism also clarifies how the operators from Chapter 5 behave through an interruption. Suppose a target distinction x and a relation $\langle x, y\rangle_\kappa$ are admissible before distraction. Successful transport must preserve not merely the component states but their operator-relevant status.

REFUSE is preserved when an excluded distractor remains outside current admissibility throughout the relevant transport:

$$x_{\text{dist}} \notin A_t$$

over the interval in which its intrusion would alter the task. A holonomy that returns the target content correctly but changes the admissibility boundary so that the distractor becomes operative may still constitute task failure.

BIND is preserved when transport retains the relation defining the joint distinction:

$$H_C^{(\mathcal{T})}\langle x, y\rangle_\kappa \sim_{\mathcal{T}} \langle x, y\rangle_\kappa.$$

Transport that preserves x and y separately while exchanging or dissolving κ lies outside the stabilizer of a conjunction-sensitive task:

$$H_C^{(\mathcal{T})} \notin K_{\mathcal{T}_{\text{bind}}}.$$

COLLAPSE is successful when post-interruption dynamics still resolve into a continuation within the task-preserving class:

$$\Gamma(A_{t_r}) \xrightarrow{\text{COLLAPSE}} \gamma^*, \qquad \gamma^* \in [\widehat{\gamma}^{\,0}]_{\mathcal{T}}.$$

Continuation failure is not itself COLLAPSE. It occurs when the transported admissibility structure no longer supports a resolution into the required equivalence class, or when the realized path falls outside it.

These relations show why endpoint accuracy can conceal different failures. The system may preserve content but lose binding, preserve binding but admit a distractor, or preserve all relevant distinctions until an incorrect final resolution. A sufficiently rich task map and transport analysis can distinguish these cases instead of compressing them into a single error label.

When the Formalism Would Be Idle

The gauge construction earns its place only if it supplies empirical or explanatory leverage unavailable from a simpler state-space account. Several outcomes would show that it does not.

If one fixed decoder applied across all task epochs predicts behavior as well as any transported decoder, then no nontrivial connection is needed for the phenomenon studied. If representational transformations cannot be estimated reliably on held-out trials, the proposed fibers and transport law lack empirical stability. If the inferred holonomy predicts no future distinction beyond endpoint distance, it adds notation without information. If rotational wave geometry contributes nothing after recurrent state is measured adequately, the physical continuation-field hypothesis does not follow from the formal transport structure. Finally, if arbitrary analytical choices about latent coordinates change whether a trial exhibits task-preserving holonomy, the claimed gauge invariance has not been achieved.

The formalism would earn its place if a connection estimated independently from ordinary task transitions predicts how representations transform through novel interruptions; if loop residuals predict specific future errors that endpoint similarity misses; if those predictions survive admissible changes of neural coordinates; and if manipulation of traveling-wave geometry selectively changes the residual transformation while preserving the component activity required by competing recurrent accounts.

This hierarchy preserves the distinction between three claims. Task-relative equivalence may be useful without gauge theory. A transport connection may be useful without wave geometry being causal. Wave geometry may be causally involved without interference being irreducibly computational. None of these implications runs automatically in reverse.

Persistence as a Gauge-Invariant Capacity

The phrase "gauge of persistence" can now be given a precise meaning. A cognitive content persists not because one neural pattern remains fixed, but because transformations among its changing realizations preserve the future distinctions that define its continuation. The permissible transformations form the task-preserving subgroup $K_{\mathcal{T}}$. The connection specifies how representations are compared across contexts. Holonomy measures the residual accumulated around an interruption-relative loop. Persistence requires that this residual act trivially on the relevant distinctions, not that it equal the identity in every neural coordinate:

$$H_C^{(\mathcal{T})} \in K_{\mathcal{T}}, \qquad \text{not necessarily} \qquad H_C^{(\mathcal{T})} = I.$$

This is stronger than saying that different neural patterns can encode the same thing. It says that the relation among those patterns may itself have a geometry, and that successful continuation depends upon the transport law preserving an equivalence class across perturbation. It is weaker than claiming that the brain has been shown to implement a literal gauge field. The state space, local regularity, transformation group, connection, and invariant observables all remain objects that must be recovered from data and validated independently.

The rotating-wave result supplies a compelling candidate physical process because it exhibits organized transport-like geometry and because its completeness predicts behavioral recovery. It does not yet establish that the observed rotation implements the connection, that its completion produces task-preserving holonomy, or that its geometry carries computational information irreducible to recurrent dynamics. Those remain the decisive empirical questions.

The next chapter draws the consequence that can already be defended without resolving them. Memory of an interrupted thought cannot be assigned exclusively to the persistent disposition structure $\mathcal{S}$, because disposition alone does not specify the recovered trajectory. Nor can it be assigned exclusively to the transient geometry, because recovery would be impossible without a stored repertoire to constrain what counts as the same content. The operative memory is distributed

between disposition and recoverable organization. Part III concludes by making that division explicit and by identifying which parts of it the present evidence measures, which it merely suggests, and which remain unknown.

Chapter 12

Memory Distributed Between Disposition and Geometry

Part III has moved from an interruption paradigm, through a formal notion of task-relative recovery, into a full transport calculus for asking what survives that interruption and what does not. This closing chapter asks a simpler question of the material assembled so far: what does any of it say about where memory of an interrupted thought actually resides. The answer this chapter defends is conditional rather than categorical, and it is worth stating the conditional form up front so the rest of the chapter cannot be mistaken for overreaching beyond it. The evidence supports treating memory as jointly dependent on persistent disposition and on dynamically recoverable organization. It does not yet establish any measurable ratio between the two, and it does not prove that wave geometry specifically is an independently stored component of memory, as opposed to a signature generated by recurrent dynamics that are themselves doing the storing. The recurrent-network competitor from Chapter 3 has not been dislodged by anything in this part of the book, and this chapter does not pretend otherwise.

Part of what makes memory talk slippery in this context is that the word is being asked to do three different jobs at once, and the three are worth pulling apart even at the cost of some initial awkwardness. Dispositional memory is the capacity retained in $P(\mathcal{S})$, the persistent repertoire a population's connectivity supports, independent of whether any particular content within that repertoire is currently active. Operative memory is content presently admissible or expressed, tracked by A_t and E_t, the quantities Chapter 5 defined for exactly this purpose. Reconstructive memory is a different capacity again: the ability to return, after a perturbation, to the task-appropriate equivalence class defined in Chapter 10. These three are causally related in an obvious sense, since reconstructive memory presumably draws on dispositional memory and produces operative memory as its outcome, but they are not the same quantity, and collapsing them produces confusion of exactly the kind this book has been trying to avoid throughout:

$$\mathsf{M}_{\mathrm{disp}}(x) \neq \mathsf{M}_{\mathrm{op}}(x, t) \neq \mathsf{M}_{\mathrm{rec}}(x, \mathcal{T}).$$

Keeping the three separate has an immediate payoff: it makes room for dissociations that a single undifferentiated notion of memory would obscure. A content can remain dispositionally available while being temporarily inadmissible, present in $P(\mathcal{S})$ but excluded from A_t by whatever process Chapter 5's REFUSE describes. A content can be expressed and yet fragile, present in E_t at one moment but poorly protected against the next distraction. And a content can disappear from current decoding entirely, undetectable in E_t or even in A_t, while remaining reconstructible from the system's ongoing dynamics, a case in which reconstructive memory survives even though

operative memory has momentarily vanished. This third case matters more than it might first appear to, because it blocks a specific and easy mistake: treating a temporary decoding failure, a trial on which a content cannot presently be read out of recorded activity, as equivalent to erasure of the content from the system altogether. The two are not the same claim, and only the second one concerns $\mathcal{S}$.

Reconstructive memory can be given a definition that makes its dependence on both disposition and dynamics explicit rather than assumed:

$$\mathsf{M}_{\mathrm{rec}}(x, \mathcal{T}) = \Pr\Big(\gamma^d \in [\widehat{\gamma}^{\,0}]_{\mathcal{T}} \mid x \in P(\mathcal{S}),\, z^{-},\, D,\, \mathcal{D}\Big),$$

where D denotes the perturbation the system undergoes and $\mathcal{D}$ denotes whatever dynamical organization governs the system's response to it. The field hypothesis this book has been developing proposes that $\mathcal{D}$ includes G_t^{wave}, the wave geometry examined at length in Chapters 6 through 11. That inclusion is a hypothesis about the contents of $\mathcal{D}$, not something the definition of $\mathsf{M}_{\mathrm{rec}}$ grants by construction. A theory in which $\mathcal{D}$ is exhausted by recurrent network state, with no independent contribution from wave geometry, is written in exactly the same formal terms and remains a live competitor.

It is worth being explicit about what "distributed" means in the phrase this chapter's title uses, because the word invites a misleading picture. It does not mean that some fraction of a memory is materially stored in synaptic weights and some other fraction is stored inside a traveling wave, as though the two were separate containers each holding a portion of the same content. Synapses are not a partial warehouse and the field is not the rest of the warehouse. What the evidence supports is a functional and counterfactual claim: persistent structure and transient organization play different roles in determining whether a content survives an interruption, and neither role is well described by the other's vocabulary. Roughly, $\mathcal{S}$ constrains what can be restored across contexts, setting the outer boundary of the repertoire available to any recovery process whatsoever, while $\mathcal{D}_t$ constrains what is restored on this particular occasion, given that boundary. Distribution, in this sense, is a claim about the division of functional labor between a constraint and a selection operating on it, not a claim about two physical storage sites sharing custody of one memory trace.

This functional framing suggests a test more direct than anything attempted so far in this part of the book, even if the test is easier to state than to run. A selective perturbation of disposition, of the kind produced by structural synaptic manipulation, should change which contents are recoverable at all, across a range of contexts, since it alters $P(\mathcal{S})$ itself. A selective perturbation of recovery geometry, holding the underlying repertoire fixed, should instead change whether an already-available content is successfully reinstated on a given trial, without changing what the system is capable of recovering in principle. If both kinds of perturbation are performed and their effects do not simply add,

$$\Delta_{\mathcal{S},\mathcal{D}} \neq \Delta_{\mathcal{S}} + \Delta_{\mathcal{D}},$$

this interaction would support genuine joint dependence between disposition and dynamics, rather than two separable contributions that happen to be measured together. No experiment reviewed in this book has performed this comparison, and it is worth acknowledging plainly how difficult it would be to perform. Manipulating disposition selectively, without also disturbing the dynamics that operate on it, and manipulating recovery geometry selectively, without touching the repertoire it draws on, are both demanding interventions, and the interaction term above is correspondingly a standard this book can state but not yet claim to have met.

Part III closes, then, by returning to the same evidential tiers that have organized the argument from Chapter 4 onward, applied now to the question of memory itself rather than to any single

operator. The empirical literature reviewed in this part supports the existence of persistent representational structure, of behaviorally predictive post-distraction dynamics, and of task-relative reconstruction as a real and measurable phenomenon rather than a theoretical fiction. It suggests, without demonstrating, that rotating waves specifically contribute to that reconstruction, since the geometry-selective interventions that would move this claim from tier 1 toward tier 2 have not yet been performed for recovery any more than they were performed for the operators examined in Part II. It does not establish, and has not come close to establishing, that wave interference performs irreducible computation in the process of reconstruction, which remains exactly the open question Chapter 8 named and this part of the book has repeatedly deferred rather than resolved.

This leaves Part IV a clean question to take up next, distinct from anything Part III has asked. Distraction tests whether a local, task-specific continuation can be repaired once interrupted, and everything examined so far has concerned that local repair. Anesthesia tests something categorically different: not whether one thread of ongoing cognition can be restored after a brief interruption, but whether the system retains the global capacity to sustain any organized continuation at all. That is a question about the regime as a whole rather than about a single content's recoverability within it, and it is where the book turns next.

Part IV

Consciousness as a Dynamical Regime

Chapter 13

Three Agents, Three Targets, No Shared Molecule

Part III examined a local failure and repair problem. A distractor perturbs an ongoing cognitive trajectory, after which the system may or may not reconstruct a task-equivalent continuation. General anesthesia presents a more extensive perturbation. The question is no longer whether one remembered item survives an interruption, but whether the nervous system retains the large-scale dynamical capacity required for contents, rules, sensations, and responses to participate in any unified continuation at all.

The three anesthetic agents considered in this part are useful precisely because they do not begin from the same molecular mechanism. Propofol, ketamine, and dexmedetomidine act primarily on different receptor systems and produce states that are neither pharmacologically nor phenomenologically identical. If they nevertheless converge upon a common alteration of large-scale neural organization as behavioral consciousness is lost, that convergence supplies evidence that the shared dynamical regime matters across mechanisms.

It does not establish that dynamics replace molecular explanation. Every dynamical change in the brain is produced through molecular, cellular, and circuit-level processes. The inferential question is instead one of explanatory level. If distinct interventions enter the nervous system through different molecular routes but repeatedly approach a similar macroscopic failure regime, then that regime may identify a common condition of consciousness more directly than any one receptor target does.

Propofol and Enhanced Inhibition

Propofol is commonly characterized by its action on $GABA_A$ receptors. By potentiating inhibitory neurotransmission, it changes the balance between excitation and inhibition across neural circuits and can produce rapid loss of behavioral responsiveness. At an elementary level, this might suggest a simple silencing account: consciousness disappears because cortical activity is suppressed.

That description is incomplete. Propofol does not turn the brain uniformly off. Neural activity continues, oscillations remain prominent, and different cortical and subcortical regions are affected in different ways. The important changes include altered coordination, changes in effective connectivity, and a reduction in the stability with which cortical population activity returns toward its ordinary operating regime after perturbation. The relevant contrast is therefore not between an electrically active conscious brain and an electrically silent unconscious one. It is between differently organized forms of activity.

In the notation developed earlier, propofol changes more than the momentary expressed set E_t. It alters the dynamics by which admissible states are maintained and by which perturbations are absorbed:

$$\dot{z}_t = F_{\mathcal{S}}^{(\text{prop})}\left(z_t, u_t; q_t\right).$$

The persistent disposition structure $\mathcal{S}$ is not erased when the agent takes effect. Learned representations and synaptic organization remain materially present. What changes is the regime in which those dispositions can become mutually available, expressed, bound, and continued.

This makes propofol relevant to the distinction between dispositional and operative memory established in Chapter 12. A person does not ordinarily lose a lifetime of stored knowledge when anesthetized and relearn it upon waking. The repertoire remains broadly preserved while its present availability and integration are profoundly altered. Whatever unconsciousness consists in, it cannot be described adequately as deletion of the stored representational space.

Ketamine and Disrupted Excitation

Ketamine begins from a different molecular target. Its best-known action is antagonism of the NMDA class of glutamate receptor, a receptor central to excitatory signaling, synaptic plasticity, and recurrent cortical interaction. Where propofol strengthens a major inhibitory system, ketamine blocks an important component of excitatory transmission.

The state ketamine produces also differs from conventional propofol anesthesia. Depending on dose and context, ketamine may produce dissociation, analgesia, vivid internally generated experience, altered bodily awareness, and varying degrees of behavioral unresponsiveness. A subject who does not respond externally under ketamine cannot therefore be assumed, solely from that absence of response, to lack all subjective experience. This distinction will matter throughout Part IV, since the experiments considered here measure neural dynamics and behavioral state more directly than they measure phenomenology.

Ketamine consequently complicates any simple equation between unresponsiveness and unconsciousness, since behavioral access, environmental responsiveness, memory formation, and phenomenal experience can dissociate from one another under it. The phrase "loss of consciousness" must therefore be indexed to the operational measure used:

$$\mathsf{C}_{\text{behavioral}}, \qquad \mathsf{C}_{\text{reportable}}, \qquad \mathsf{C}_{\text{phenomenal}}.$$

An experiment may establish a transition in the first without directly settling the status of the third, which is precisely what makes this agent especially informative rather than less useful: if a dissociative state produced through NMDA-receptor antagonism converges dynamically with a GABAergic anesthetic on some measures while diverging on others, the pattern can help separate dynamics associated with behavioral disconnection from those associated with complete absence of experience. Convergence should not be assumed globally. It must be specified variable by variable.

Dexmedetomidine and Suppressed Arousal

Dexmedetomidine enters through a third molecular route. It is an agonist at α_2-adrenergic receptors and acts strongly upon systems involved in arousal, including pathways associated with the locus coeruleus. Its sedative state is often described as resembling aspects of non-rapid-eye-movement sleep more closely than the states produced by propofol or ketamine. Subjects can sometimes be roused relatively readily despite appearing deeply sedated before stimulation.

The agent therefore reaches behavioral unresponsiveness partly by changing arousal regulation rather than primarily by potentiating GABAergic inhibition or blocking NMDA-mediated excitation. Once again, the downstream effects are distributed. Altering a subcortical arousal system changes cortical responsiveness, phase coordination, population stability, and the conditions under which distant regions can participate in a shared evolving state.

Dexmedetomidine is particularly important for the continuation-field hypothesis because it demonstrates that a global change in cognitive availability need not begin with a molecular intervention directed chiefly at local cortical computation. A change in arousal can reorganize the boundary conditions under which cortical distinctions become admissible. In the language of Chapter 5, the agent need not erase elements from $P(\mathcal{S})$ in order to produce a radical contraction or fragmentation of A_t.

The sleep-like description must nevertheless be handled cautiously. Pharmacological sedation is not ordinary sleep, and resemblance in selected oscillatory or behavioral features does not establish identity of mechanism or subjective state. The comparison identifies a possible shared arousal pathway, not an equivalence between naturally sleeping and pharmacologically sedated brains.

Different Entry Points, Overlapping Downstream Systems

The molecular diversity of these three agents supplies a real inferential advantage, but it should not be read too literally, as though each acted upon a sealed and independent brain. Their receptor systems participate in densely coupled networks, and their downstream consequences can overlap substantially. Enhanced inhibition alters excitation. NMDA antagonism can disinhibit some populations while suppressing others. Adrenergic modulation changes cortical gain and responsiveness. At the circuit level, distinct molecular interventions can meet long before they produce a measured behavioral endpoint.

The relevant architecture is therefore many-to-one:

$$\begin{aligned} \text{propofol} &\longrightarrow \text{GABAergic modulation} \longrightarrow \mathcal{N}_{\text{prop}}, \\ \text{ketamine} &\longrightarrow \text{NMDA antagonism} \longrightarrow \mathcal{N}_{\text{ket}}, \\ \text{dexmedetomidine} &\longrightarrow \alpha_2\text{-adrenergic modulation} \longrightarrow \mathcal{N}_{\text{dex}}, \end{aligned}$$

where the downstream network regimes

$$\mathcal{N}_{\text{prop}}, \quad \mathcal{N}_{\text{ket}}, \quad \mathcal{N}_{\text{dex}}$$

may overlap in some dynamical dimensions while remaining different in others.

The convergence question is not whether

$$\mathcal{N}_{\text{prop}} = \mathcal{N}_{\text{ket}} = \mathcal{N}_{\text{dex}}.$$

That would be implausibly strong. It is whether there exists a dynamical property $\mathcal{D}^*$ such that

$$\mathcal{D}^* \subseteq \mathcal{N}_{\text{prop}} \cap \mathcal{N}_{\text{ket}} \cap \mathcal{N}_{\text{dex}},$$

and whether alteration of $\mathcal{D}^*$ reliably accompanies the relevant transition in behavioral or conscious state.

The studies considered in the next chapter identify two candidates for such a shared property. The first is dynamical destabilization: a reduction in the capacity of cortical population activity to absorb perturbation and return toward its ordinary operating range. The second is a convergent alteration of phase alignment, including reduced alignment between neighboring prefrontal areas and increased alignment between corresponding regions across the hemispheres under ketamine and dexmedetomidine. These findings concern organization rather than overall quantity of activity.

Why Convergence Is Informative

Suppose three interventions act through distinct proximal mechanisms,

$$M_1, \quad M_2, \quad M_3,$$

and all produce a behavioral transition B. If each also produces a shared dynamical alteration D, then the observed structure is

$$M_i \longrightarrow D \longrightarrow B, \qquad i \in \{1, 2, 3\},$$

or at least a pattern consistent with that causal chain.

The diversity of M_i weakens the possibility that the association between D and B is peculiar to one receptor system. It does not eliminate confounding, establish mediation, or prove that D is necessary. The agents may share an unmeasured downstream mechanism that independently produces both D and B:

$$M_i \longrightarrow L \longrightarrow \begin{cases} D, \\ B. \end{cases}$$

Alternatively, D may be a consequence rather than a cause of the behavioral transition:

$$M_i \longrightarrow B \longrightarrow D.$$

A third possibility is that different causal pathways produce superficially similar values of a coarse dynamical measure without producing the same underlying neural regime.

Convergence therefore strengthens an inference only to the resolution at which convergence has actually been demonstrated. Three agents can all reduce a stability statistic while differing in which circuits become unstable, which contents remain available, and whether subjective experience continues. The shared statistic identifies a candidate common dimension; it does not prove identity of state.

The logic becomes stronger when several conditions are met. The dynamical change should occur near or before the behavioral transition rather than only afterward. Its magnitude should track depth of impairment within each condition. Recovery of the dynamical property should accompany recovery of responsiveness. The relationship should survive controls for dose, movement, sensory input, and global changes in firing or spectral power. Most importantly, selective intervention on the proposed dynamical property should change the behavioral state without merely reproducing the agent's molecular action.

The existing evidence does not satisfy all of these conditions. It does enough to make convergence scientifically important and insufficient to turn convergence into demonstrated mechanism.

Organization Rather Than Location

The three-agent comparison shifts the explanatory question away from the search for a single molecular switch for consciousness. Propofol, ketamine, and dexmedetomidine do not reveal one receptor whose state simply determines whether experience exists. Their diversity suggests that different molecular routes can alter a shared capacity of large-scale neural organization.

This is compatible with a regime account. On such an account, consciousness depends upon the nervous system occupying a range of dynamical conditions within which perturbations can be

absorbed, representations can remain mutually accessible, and activity in one region can become operative context for activity elsewhere. The regime is multiply realizable at the molecular level:

$$\{M_1, M_2, M_3, \ldots\} \longrightarrow \mathcal{R}_{\mathsf{C}}.$$

Distinct molecular arrangements may support the conscious regime, and distinct molecular interventions may disrupt it.

Multiple realizability does not make molecular details irrelevant. The molecular route determines which components of the regime fail first, which residual experiences remain possible, how readily the state can be reversed, and what physiological risks accompany the transition. A regime-level description complements rather than replaces lower-level pharmacology.

Nor does a regime account locate consciousness in a global average. Large-scale integration can be lost because specific directional relations fail, because the system becomes excessively synchronized, because neighboring regions decouple, because long-range phase structure reorganizes, or because local dynamics cease to remain within a recoverable range. "Global organization" names a family of possibilities, not a sufficient mechanism.

From Local Repair to Global Capacity

The distraction paradigm in Part III asks whether a task-relevant trajectory can be restored after a localized perturbation. Anesthesia asks whether the system remains in a regime capable of supporting such restoration at all. The relationship can be written as a difference of scale:

$$\text{distraction}: \quad \gamma^- \longrightarrow D \longrightarrow \gamma^+,$$

where recovery is possible if

$$\gamma^+ \in [\widehat{\gamma}^{\,0}]_{\mathcal{T}},$$

whereas anesthesia may alter the dynamics so extensively that the relevant task-equivalence class is no longer reachable:

$$[\widehat{\gamma}^{\,0}]_{\mathcal{T}} \cap \mathrm{Reach}_{F^{(\mathrm{anesth})}}(z_t) = \varnothing.$$

This expression should not be interpreted as a claim already established by the anesthesia data. It states the continuation-field hypothesis in a form that can fail. The hypothesis is that anesthetic-induced unresponsiveness results partly from alteration of the dynamics governing reachability, transport, and recovery, rather than from deletion of the persistent representations themselves.

A weaker version requires no field-specific mechanism. Ordinary recurrent-network dynamics can also become unstable, fragmented, or trapped in regimes that no longer support global coordination. The field account becomes distinctive only if wave geometry contributes predictive or causal information beyond that recurrent description. Part IV therefore inherits the same three-tier evidential hierarchy used in Part II:

tier 1: dynamics predict anesthetic state,

tier 2: dynamics causally regulate the transition,

tier 3: field geometry performs indispensable organization.

The fact that three agents converge upon a dynamical alteration can strengthen tier 1 substantially. It does not, by itself, establish either of the higher tiers.

What the Comparison Can Establish

The three-agent design can support a limited but important inference. If molecularly distinct interventions produce a common alteration in neural stability or phase organization at the transition to behavioral unresponsiveness, then no explanation restricted to the unique proximal target of one agent can be sufficient for the shared outcome. Some downstream organization must enter the explanation.

That conclusion is weaker than saying that the shared dynamical alteration causes unconsciousness. It is also stronger than saying merely that anesthetics affect the brain. It identifies a level at which heterogeneous molecular perturbations become comparable:

$$\text{different proximal mechanisms} \longrightarrow \text{partly convergent dynamical regimes.}$$

The comparison does not show that all three agents abolish phenomenal experience in the same way. It does not show that every property they share is relevant to consciousness. It does not show that a single dynamical variable is sufficient for a conscious state. It does not establish that electric fields, rather than recurrent circuits, generate the shared organization. It does not address why any organized neural activity should be accompanied by experience.

What it does is create an empirical opportunity. Molecular diversity acts as a kind of natural perturbational basis. Features that remain specific to one agent are less plausible as universal explanations of the shared behavioral transition. Features that converge across agents become candidates for a common regime-level condition. The next chapter examines the strongest candidates reported so far: destabilization of cortical dynamics and convergent changes in phase alignment. Its question is not whether these observations solve consciousness, but whether they identify a shared loss of organized continuation across otherwise different pharmacological routes.

Chapter 14

Convergent Destabilization Across Mechanism

The preceding chapter established why propofol, ketamine, and dexmedetomidine constitute an informative comparison. They enter the nervous system through different proximal molecular pathways and produce importantly different anesthetic states. Their value for the present argument lies in the possibility that these distinct interventions nevertheless converge upon a shared alteration of large-scale neural organization. This chapter examines two reported forms of convergence: destabilization of cortical population dynamics and reorganization of phase alignment across cortical regions.

The strongest conclusion available from these findings is not that a single neural signature explains consciousness. It is that behavioral unresponsiveness produced through different molecular mechanisms is accompanied by common changes in how cortical activity maintains, coordinates, and restores itself. This shifts the explanatory target from the presence of neural activity to the regime governing its evolution.

Stability as Return After Perturbation

A dynamical system is stable, in the relevant sense, when sufficiently small perturbations do not drive it indefinitely away from its ordinary operating regime. The system need not return to an identical microstate. It must remain within, or return toward, a bounded region in which its characteristic functions remain available.

Let the unperturbed cortical dynamics be written

$$\dot{z}_t = F(z_t, u_t),$$

and let $z^*(t)$ denote a reference trajectory within the ordinary waking regime. A perturbation produces

$$z_t = z^*(t) + \delta z_t.$$

Local recovery requires the displacement to remain bounded and, under stronger conditions, to contract:

$$\|\delta z_{t+\Delta t}\| < \|\delta z_t\|.$$

This elementary expression is only a starting point. Neural activity is nonstationary, high-dimensional, and continually driven by internal and external input. The relevant reference object may therefore

be an attractor, a metastable manifold, a distribution of trajectories, or a time-dependent operating region rather than a fixed point.

A more suitable condition compares the evolving state with a viable set $\mathcal{V}$:

$$d(z_{t+\Delta t}, \mathcal{V}) < d(z_t, \mathcal{V})$$

after an experimentally or naturally occurring perturbation. The set $\mathcal{V}$ contains states compatible with the system's ordinary organized regime. Stability then means not immobility but recoverability.

This definition connects directly to Part III. Recovery from distraction was treated there as return to a task-equivalent continuation rather than return to an identical state. Dynamical stability generalizes that logic. A stable conscious regime may tolerate substantial local displacement so long as the system retains a route back into a region supporting coherent continuation. Instability means that comparable perturbations produce larger, less predictable, or less recoverable departures.

A Shared Loss of Dynamical Stability

Eisen and colleagues reported in 2026 [7] that propofol, ketamine, and dexmedetomidine produce a similar destabilization of neural population dynamics as animals lose behavioral responsiveness. The importance of the result lies in the diversity of the molecular routes involved. Enhanced GABAergic inhibition, NMDA-receptor antagonism, and altered adrenergic arousal regulation do not begin from the same proximal mechanism, yet each condition was associated with a reduction in the cortex's capacity to absorb perturbation and remain within its ordinary dynamical range.

The shared observation can be represented as

$$M_i \longrightarrow F_i^{(\text{anesth})}, \qquad i \in \{\text{prop}, \text{ket}, \text{dex}\},$$

with

$$\mathfrak{S}\left(F_i^{(\text{anesth})}\right) < \mathfrak{S}\left(F^{(\text{awake})}\right),$$

where $\mathfrak{S}$ denotes an empirically defined stability measure. The molecular interventions differ, and the resulting cortical dynamics need not be identical, but the direction of change in stability is shared.

This is not simply a reduction in activity. A system can exhibit low average activity while remaining stable, and it can exhibit intense activity while being dynamically unstable. Nor is stability equivalent to rigidity. A system that cannot leave one narrow state may be locally resistant to perturbation while lacking the flexibility required for cognition. The relevant property is bounded flexibility: the capacity to move through a repertoire of states without losing the organization that allows those states to remain part of one viable regime.

One may distinguish at least three possibilities:

rigidity:	few accessible states and little adaptive movement,
organized flexibility:	many accessible states with recoverable transitions,
destabilization:	perturbations produce poorly bounded or poorly recoverable motion.

Conscious waking activity is hypothesized to occupy the middle condition rather than merely maximizing or minimizing variability.

The reported convergence therefore points toward a failure of regulation rather than a uniform direction of neural amplitude. Under anesthesia, cortical trajectories may continue to evolve, but their evolution no longer exhibits the same capacity to preserve or restore the relations characteristic of the waking regime.

Stability Is Scale-Relative

A stability claim is incomplete until the relevant scale and variables are specified. A population may be stable in mean firing rate while unstable in latent-state geometry. Local activity may remain bounded while coordination between regions deteriorates. A global signal may appear regular precisely because the underlying system has become excessively constrained.

Let

$$\mathfrak{S}_{\ell,\tau,\Omega}$$

denote stability measured at spatial scale ℓ, temporal scale τ, and through observation map Ω. Then

$$\mathfrak{S}_{\ell_1,\tau_1,\Omega_1} \neq \mathfrak{S}_{\ell_2,\tau_2,\Omega_2}$$

need not indicate a contradiction. The quantities may describe different levels of the same system.

This qualification is especially important for comparisons among anesthetic conditions. Propofol may produce one pattern of local regularity and long-range disruption, while ketamine produces another. A common decrease in a selected stability statistic does not imply that every dynamical scale changes in the same way. The shared statistic identifies a candidate dimension of convergence; it does not collapse the three conditions into a single neural state.

The reference regime also matters. If the waking baseline is itself heterogeneous, distance from a single average waking trajectory may exaggerate instability. A stronger analysis compares anesthetized activity with a distribution of viable waking dynamics:

$$p_{\text{wake}}(\gamma),$$

and asks whether perturbations under anesthesia reduce the probability of returning to its task-relevant support:

$$\Pr(\gamma_{t+\Delta t} \in \operatorname{supp}_{\mathcal{T}} p_{\text{wake}} \mid D)\,.$$

The continuation-field interpretation concerns precisely this capacity for return, not generic resemblance to an average waking pattern.

Phase Alignment Across Cortical Distance

A second form of convergence concerns phase relations. Bardon and colleagues reported in 2025 [2] that ketamine and dexmedetomidine altered cortical phase alignment in similar directions. Alignment decreased between neighboring prefrontal areas while increasing between corresponding regions across the two hemispheres.

This pattern is more informative than a general statement that synchrony changed. Synchronization is not uniformly beneficial, and desynchronization is not uniformly destructive. What matters is which regions align, at which frequencies, with what directionality, and under what task or behavioral condition. Increased alignment can reflect effective coordination, but it can also reflect a loss of differentiated local dynamics. Decreased alignment can permit functional specialization, but it can also sever relations required for information exchange.

Let $\phi_i(t)$ denote the phase of an oscillatory component in region i. A simple pairwise alignment statistic is

$$R_{ij} = \left| \frac{1}{N} \sum_{n=1}^{N} e^{\,\mathrm{i}(\phi_i(t_n) - \phi_j(t_n))} \right|.$$

Values near one indicate a consistent phase relationship across observations; values near zero indicate variable relative phase. The reported pattern can be written schematically as

$$\Delta R_{\text{neighbor}} < 0, \qquad \Delta R_{\text{bilateral}} > 0.$$

The coexistence of these two changes suggests reorganization rather than indiscriminate loss of synchrony. Local neighboring relations become less consistently aligned while homologous cross-hemispheric relations become more consistently aligned. The cortical system has not simply become less coordinated. Its topology of coordination has changed.

A useful network representation assigns each cortical region a node and each phase relationship a weighted edge:

$$W_{ij} = R_{ij}.$$

Anesthetic transition then produces

$$W^{(\text{awake})} \longmapsto W^{(\text{anesth})}.$$

The relevant comparison concerns the redistribution of weights across spatial classes:

$$\Delta W = W^{(\text{anesth})} - W^{(\text{awake})}.$$

A shared redistribution across molecularly different conditions is evidence that large-scale organization is being driven toward a partly common regime.

Integration Is Not Maximum Synchrony

The phase-alignment result warns against defining consciousness as global synchrony. If consciousness required only more alignment, increased bilateral alignment under anesthesia would point in the wrong direction. The finding instead supports a differentiated notion of integration.

A viable integrative regime must coordinate distributed regions while preserving distinctions among them. If every region follows the same phase pattern, the system may be globally coherent but informationally impoverished. If every region evolves independently, the system may preserve local differentiation but lack the relations needed for unified continuation. The relevant regime lies between these extremes.

Let $\mathcal{I}$ denote integration and $\mathcal{D}$ differentiation. A schematic viability functional may be written

$$\mathcal{V}_{\text{int}} = \Phi(\mathcal{I}, \mathcal{D}),$$

with

$$\frac{\partial \Phi}{\partial \mathcal{I}} > 0$$

only where sufficient differentiation remains, and

$$\frac{\partial \Phi}{\partial \mathcal{D}} > 0$$

only where sufficient coordination remains. Neither variable is beneficial without qualification.

The observed combination of reduced neighboring alignment and increased bilateral alignment may reflect a change in this balance. Local relations needed for fine-grained coordination weaken, while more distant homologous regions become locked into a coarser shared rhythm. The result may be simultaneously more synchronized along one dimension and less capable of differentiated integration overall.

This interpretation remains provisional. Pairwise phase alignment does not directly measure information exchange, causal influence, binding, or phenomenological unity. It identifies a structured change whose functional significance must be tested rather than inferred from synchrony alone.

From Phase Pattern to Admissibility

The continuation-field hypothesis interprets phase organization as a candidate component of the conditions governing which neural events can participate in a shared trajectory. If neighboring regions lose the phase relations needed for their activity to become mutually effective, locally represented contents may cease to function as admissible context for one another. If bilateral regions become excessively aligned, distinct contributions may become less independently available.

In the notation of Chapter 5, the proposed relation is

$$G_t^{\text{phase}} \longrightarrow A_t \longrightarrow E_t,$$

where G_t^{phase} denotes the spatial organization of phase relations. This chain is a hypothesis, not a result established by the alignment statistic itself.

Tier 1 requires that phase topology predict the anesthetic or behavioral state after relevant controls:

$$I\Big(G_t^{\text{phase}}; B_t \mid C_t\Big) > 0.$$

The reported convergence supports this kind of claim, particularly where the spatial pattern discriminates responsive from unresponsive conditions.

Tier 2 would require evidence that changing the phase topology alters integrative capacity:

$$p\Big(\mathcal{I}_t \mid \text{do}(G_t^{\text{phase}} = g_1), C_t\Big) \neq p\Big(\mathcal{I}_t \mid \text{do}(G_t^{\text{phase}} = g_2), C_t\Big).$$

Tier 3 would require the stronger demonstration that the geometry of phase relations performs a task-relevant transformation not captured by recurrent population state:

$$I\Big(\mathcal{O}_t; G_t^{\text{phase}} \mid R_t^{\text{recurrent}}, u_t, C_t\Big) > 0,$$

combined with a geometry-selective intervention.

The convergence evidence remains principally at tier 1. It shows that organization changes systematically across anesthetic transitions and that partly similar changes arise through different molecular routes. It does not yet show that phase alignment itself performs the lost integrative work.

Two Forms of Convergence

Dynamical destabilization and phase reorganization are related but not identical findings. Stability concerns how trajectories respond to displacement. Phase organization concerns the spatial relations among oscillatory processes. One may contribute to the other:

$$G_t^{\text{phase}} \longrightarrow \mathfrak{S}_t,$$

or both may emerge from an unmeasured change in recurrent circuitry:

$$L_t \longrightarrow \begin{cases} G_t^{\text{phase}}, \\ \mathfrak{S}_t. \end{cases}$$

A third possibility is reciprocal coupling:

$$G_t^{\text{phase}} \rightleftarrows \mathfrak{S}_t.$$

These alternatives cannot be distinguished by observing that both quantities change near loss of responsiveness. Their temporal ordering, conditional dependence, and response to selective intervention must be measured.

The continuation-field account predicts a particular relationship. A disruption in spatial phase organization should alter the transport relations by which cortical activity is maintained within a viable regime. This should increase the holonomy defect introduced in Chapter 11 and reduce the probability of task-relative return after perturbation:

$$\delta_{\mathrm{hol}}^{(\mathcal{T})} \uparrow \quad \Longrightarrow \quad R_{\mathcal{T}} \downarrow .$$

At a global scale, repeated failures of task-relative transport would appear as reduced dynamical stability:

$$\mathbb{E}[R_{\mathcal{T}} \mid D] \downarrow \quad \Longrightarrow \quad \mathfrak{S} \downarrow .$$

This provides a bridge between Parts III and IV. The local recovery score and the global stability measure may describe the same general property at different scales: whether perturbation preserves access to a viable continuation. The bridge remains a theoretical proposal until the quantities are measured in the same preparation and shown to relate as predicted.

The Logic of Convergent Evidence

The inferential force of convergence can be expressed through competing causal models. Under a molecularly specific account, each anesthetic condition produces unresponsiveness through a largely distinct downstream mechanism:

$$M_i \longrightarrow L_i \longrightarrow B.$$

Under a shared-regime account, distinct molecular pathways converge upon a common dynamical condition:

$$M_i \longrightarrow D^* \longrightarrow B.$$

The observation that a similar D^* appears across conditions raises the posterior plausibility of the second model. It does not prove it, because a common measured consequence may lie downstream of B or share a cause with it:

$$M_i \longrightarrow L_i \longrightarrow \begin{cases} D^*, \\ B. \end{cases}$$

The evidential gain from convergence depends upon specificity. If D^* is defined vaguely as "neural activity changes," convergence is nearly guaranteed and carries little information. If it specifies a shared direction of change in perturbational stability, spatial phase topology, or return dynamics, the coincidence is more restrictive and therefore more informative.

The gain also depends upon independence. The three molecular pathways are not causally independent once they enter the same nervous system. Their downstream effects overlap, and all sufficiently deep anesthetic states may share nonspecific consequences such as reduced movement, changed sensory input, altered respiration, or modified neuromodulation. These factors must be controlled before the shared dynamical signature can be attributed specifically to loss of integrative capacity.

Convergence is therefore best treated as triangulation. Each intervention rules out some explanations that are peculiar to the others, while the remaining shared pattern identifies a candidate common mechanism. Triangulation narrows the space of explanations. It does not reduce that space to one.

Prediction, Mediation, and Necessity

Three questions should remain separate. The first is whether destabilization or phase reorganization predicts behavioral state:

$$p(B \mid D^*) \neq p(B).$$

The second is whether the dynamical change mediates part of the effect of the molecular intervention:

$$M_i \longrightarrow D^* \longrightarrow B.$$

The third is whether preservation of the dynamical property is necessary or sufficient for the relevant form of consciousness.

Prediction is the least demanding and best supported. Mediation requires temporal and causal analysis capable of separating the path through D^* from alternative pathways. Necessity would require showing that the relevant conscious capacity fails whenever the dynamical organization fails, across a sufficiently broad range of conditions. Sufficiency would require restoring the organization and thereby restoring the capacity while other relevant conditions remain impaired.

The existing studies do not establish necessity or sufficiency. They identify shared dynamical correlates and, in the case of perturbational analyses, characterize the altered response properties of the cortical system. That is substantial evidence for a regime-level account and insufficient evidence for identifying the regime with consciousness itself.

Behavioral Responsiveness and Experience

The dependent variable must also remain explicit. An animal's correct response provides evidence of perceptual access, memory, motor capacity, and task engagement. Failure to respond may result from disruption at any of these stages. Under anesthesia, immobility or absent task performance does not uniquely identify absence of subjective experience.

Let

$$B_t$$

denote behavioral responsiveness and

$$C_t^{\text{phen}}$$

denote phenomenal consciousness. The experiments directly support relationships involving B_t:

$$D^* \longleftrightarrow B_t.$$

They do not directly establish

$$D^* \longleftrightarrow C_t^{\text{phen}}.$$

This limitation is especially important for the dissociative case discussed in Chapter 13, under which behavioral disconnection can coexist with vivid internal experience. A dynamical property shared across all three agents may therefore track environmental responsiveness or externally organized continuation more closely than experience as such.

That possibility does not make the result irrelevant to consciousness. Conscious experience in ordinary waking life includes a capacity for temporally extended, environmentally responsive integration. A mechanism that supports this capacity may be an important component of consciousness without exhausting it. The evidence must be assigned to that narrower claim.

What Has Converged

The two studies considered here establish convergence at a level above proximal molecular action and below a complete theory of consciousness. Across distinct anesthetic conditions, cortical dynamics become less capable of remaining within or returning toward their ordinary operating regime. Across at least two of those conditions, phase alignment is reorganized in a similar spatial pattern, with reduced neighboring alignment and increased bilateral alignment.

These findings support the proposition that anesthetic-induced behavioral unresponsiveness involves a change in neural organization, not merely a uniform reduction in neural activity. They make a common dynamical regime a serious explanatory target. They do not establish that the shared regime is identical across conditions, that phase reorganization causes destabilization, that destabilization causes loss of experience, or that traveling fields implement an irreducible computation.

The conclusion can be stated as a constrained inference:

distinct proximal mechanisms

$$\Downarrow$$

partly convergent dynamical alterations

$$\Downarrow$$

candidate common loss of integrative capacity.

The first implication is empirically supported. The second remains an interpretation requiring further causal work.

The next chapter makes this hierarchy explicit. It separates molecular intervention, dynamical destabilization, loss of integrative capacity, behavioral unresponsiveness, and phenomenal consciousness into distinct claims rather than allowing them to collapse into one another. That separation is necessary because the empirical arrows connecting them differ sharply in how directly they can be tested and in how strongly the current evidence supports them.

Chapter 15

The Explicit Hierarchy of Claims

The preceding chapters have moved repeatedly across several explanatory levels: molecular intervention, altered cortical dynamics, reduced integrative capacity, behavioral unresponsiveness, and consciousness. These levels are related, but they are not interchangeable. A change at one level does not establish every proposed consequence farther along the chain. The purpose of this chapter is to prevent those transitions from being compressed into a single claim.

The complete hierarchy is

$$M \longrightarrow D \longrightarrow I \longrightarrow B \longrightarrow C,$$

where M denotes a molecular intervention, D a measurable alteration of neural dynamics, I a change in integrative or continuation capacity, B a change in behavioral responsiveness, and C a change in phenomenal consciousness. The five terms describe different kinds of variables. Each arrow therefore requires its own evidence.

The first arrow is pharmacological and physiological. The second is functional: it asks whether the measured dynamics support or impair integration. The third connects an inferred neural capacity to observable behavior. The fourth is the epistemic bridge from behavior to experience. Evidence becomes less direct as the chain proceeds. The fact that the first relation is well established does not confer equal confidence upon the last.

The hierarchy is not intended to imply that causation proceeds only in a single direction or through a single serial pathway. Molecular interventions affect many neural systems simultaneously. Behavior changes sensory input and internal state. Conscious experience, if present, may participate in ongoing control rather than appearing only as a final consequence. The chain isolates one explanatory route so that its component claims can be tested separately.

The First Arrow: Molecular Intervention to Neural Dynamics

The relation

$$M \longrightarrow D$$

is the most directly accessible arrow in the hierarchy. An anesthetic intervention is administered under controlled conditions, and neural dynamics are measured before, during, and after the transition. The relevant dynamical variables may include perturbational stability, phase alignment, traveling-wave geometry, spectral organization, population-state trajectories, or effective connectivity.

For a dynamical quantity D_t, the basic causal contrast is

$$\mathbb{E}[D_t \mid \mathrm{do}(M = m_1)] - \mathbb{E}[D_t \mid \mathrm{do}(M = m_0)].$$

Because the intervention is externally imposed, causal inference is stronger here than in a comparison between naturally occurring correct and error trials. Dose variation, randomized timing, within-subject comparisons, and recovery measurements can strengthen the design further.

Even this arrow is not represented adequately by a single difference score. The intervention may alter respiration, circulation, temperature, movement, sensory input, and neuromodulatory state, each of which can affect neural recordings. The causal effect on D is real at the level of the total intervention, but identifying the pathway by which it occurs requires additional control:

$$M \longrightarrow \{L_1, L_2, \ldots, L_n\} \longrightarrow D.$$

The intermediate variables L_i include both intended receptor-mediated effects and systemic consequences.

The evidence reviewed in Chapters 13 and 14 strongly supports the first arrow. The three anesthetic agents examined there alter cortical dynamical stability and phase organization. The precise downstream pathways differ, and the resulting neural states are not identical, but the direction of selected dynamical changes converges across conditions.

That conclusion is substantial and limited. It establishes that anesthetic interventions reorganize neural dynamics. It does not yet establish what those dynamics do.

The Second Arrow: Dynamics to Integrative Capacity

The relation

$$D \longrightarrow I$$

is the central mechanistic claim of Part IV. It asks whether destabilization and altered phase topology reduce the system's capacity to preserve, bind, transport, and restore task-relevant distinctions.

Integrative capacity is not directly identical to any one measured signal. It is a functional property defined through what the system remains able to do. Let $\mathcal{U}$ denote a family of perturbations or future demands, and let $\mathcal{T}$ range over a family of tasks. A schematic integrative-capacity functional is

$$I_t = \mathbb{E}_{\mathcal{T},\mathcal{U}} \left[R_{\mathcal{T}} \mid z_t, \mathcal{U} \right],$$

where $R_{\mathcal{T}}$ is the task-relative recovery score introduced in Chapter 10. A system has high integrative capacity when it can preserve or reconstruct relevant continuations across a sufficiently broad family of perturbations and task demands.

This formulation distinguishes integration from connectivity and synchrony. Two regions may be anatomically connected without their activity being mutually available in the present regime. They may be phase-aligned without preserving differentiated information. They may exchange signals without supporting recovery after perturbation. Integration is therefore defined by continuation capacity rather than by the mere existence of relations among signals.

Evidence that D predicts I has the form

$$I(D; I \mid C) > 0,$$

where C contains relevant controls. This establishes association. A causal claim requires

$$p(I \mid \mathrm{do}(D = d_1), C) \neq p(I \mid \mathrm{do}(D = d_2), C).$$

The intervention must manipulate the dynamical property rather than merely reproduce the entire anesthetic condition, since reproducing the full condition would leave the causal contribution of D unresolved.

For the continuation-field hypothesis, the intervention must be more specific still. It should alter a feature of spatial or phase geometry while preserving, as far as possible, spectral power, local firing, sensory input, and recurrent population state:

$$\mathrm{do}\left(G_t^{\mathrm{wave}} = g'\right) \longrightarrow \Delta I_t.$$

This is the tier-2 test carried forward from Chapter 8. Tier 3 would additionally require evidence that the geometric contribution is not captured by the strongest adequate recurrent-state model.

The existing evidence supports the second arrow indirectly. Destabilization is defined partly through reduced recovery after perturbation, and phase-topology changes are plausible candidates for altered coordination. But the full causal claim that these specific dynamical alterations produce loss of integrative capacity has not been established through selective intervention.

The distinction between definition and mechanism matters here. If I is defined simply as the selected stability statistic D, then

$$D \longrightarrow I$$

becomes true by stipulation. Integrative capacity must instead be measured through independently defined performance under perturbation, preservation of task-relative information, cross-context transport, or recovery of future discriminations. Only then can the relationship between dynamics and function be tested rather than assumed.

The Third Arrow: Integrative Capacity to Behavioral Responsiveness

The relation

$$I \longrightarrow B$$

connects neural organization to observable behavior. A system that cannot preserve task-relevant distinctions, coordinate perception with memory, or carry an intention into action will often fail to respond appropriately. The relationship is therefore plausible, but it is not one-to-one.

Behavioral responsiveness depends upon several capacities:

$$B = f(I, S_{\mathrm{sens}}, M_{\mathrm{motor}}, A_{\mathrm{arousal}}, V_{\mathrm{value}}, C_{\mathrm{context}}),$$

where S_{sens} denotes sensory processing, M_{motor} motor capacity, A_{arousal} arousal, V_{value} motivational relevance, and C_{context} understanding of the response context.

A failure of behavior may arise from impairment anywhere in this system. The subject may not perceive the instruction, may not retain it, may lack sufficient arousal to act, may be unable to execute the response, or may remain internally aware while disconnected from the environment. Behavioral unresponsiveness therefore supplies evidence of impaired access or control without uniquely localizing the failure to integration.

Conversely, a simple response may survive despite substantial loss of integrative capacity. Reflexive, overlearned, or locally organized behavior may require less cross-regional continuation than novel report, flexible rule use, or delayed decision-making. The choice of behavioral measure changes the strength of the inference:

$$B_{\mathrm{reflex}} \neq B_{\mathrm{command}} \neq B_{\mathrm{flexible}} \neq B_{\mathrm{report}}.$$

Preservation of one does not guarantee preservation of the others.

The third arrow should therefore be evaluated across a graded family of behaviors rather than through a single binary threshold. If declining integrative capacity is genuinely responsible for behavioral disconnection, it should predict an ordered loss of functions according to their dependence upon long-range coordination, temporal maintenance, and recovery from perturbation.

A causal mediation model would ask whether the effect of M on B is transmitted partly through D and I:

$$M \longrightarrow D \longrightarrow I \longrightarrow B.$$

The mediated effect must be distinguished from parallel pathways:

$$M \longrightarrow \begin{cases} D \longrightarrow I \longrightarrow B, \\ S_{\text{sens}} \longrightarrow B, \\ M_{\text{motor}} \longrightarrow B, \\ A_{\text{arousal}} \longrightarrow B. \end{cases}$$

The existence of these alternatives does not defeat the integrative account. It determines the controls and comparisons required to isolate it.

The Fourth Arrow: Behavior to Phenomenal Consciousness

The relation

$$B \longrightarrow C$$

is not an ordinary causal arrow of the same kind as the preceding three. Behavioral responsiveness is used as evidence about consciousness, but consciousness need not be produced by behavior. The arrow marks an inferential bridge:

$$B \rightsquigarrow C.$$

Replacing the causal arrow with $\rightsquigarrow$ makes the epistemic status explicit.

When a subject follows a novel command, reports a stimulus, or later recalls an experience, behavior provides strong evidence that some relevant conscious processing occurred. When a subject does not respond, the inference is asymmetric. Absence of behavior may indicate absence of experience, but it may also reflect sensory disconnection, motor impairment, amnesia, reduced arousal, or inability to translate experience into the required response.

Formally,

$$\Pr(C = 1 \mid B = 1)$$

may be high under a well-designed report condition, while

$$\Pr(C = 0 \mid B = 0)$$

need not be comparably high. The exact probabilities cannot be assumed, and they vary across anesthetic conditions and behavioral assays.

The dissociative agent examined in Chapter 13 makes this asymmetry especially clear, since behavioral disconnection under it may coexist with vivid internal experience. A dynamical signature correlated with unresponsiveness in that condition may therefore track loss of environmental integration, reportability, or motor access rather than absence of phenomenal content, and internally experienced but externally disconnected states of this kind make the fourth arrow especially hard to close.

The relevant variables should remain separate:

$$C^{\text{phen}}, \qquad C^{\text{access}}, \qquad C^{\text{report}}, \qquad B^{\text{response}}.$$

Phenomenal consciousness concerns whether experience is present. Access consciousness concerns whether content is available for reasoning and control. Reportability concerns whether it can be communicated. Behavioral responsiveness concerns whether an observable action occurs. These categories interact, but no pair is definitionally identical.

Part IV can make a strong claim about behavioral and integrative state without settling phenomenal consciousness:

$$D \longrightarrow I \longrightarrow B.$$

Moving from that result to

$$D \longrightarrow C^{\text{phen}}$$

requires additional assumptions or evidence. Those assumptions should be named rather than hidden in the phrase "loss of consciousness."

The Revised Chain

The hierarchy should therefore be written with different arrow types:

$$M \longrightarrow D \dashrightarrow I \dashrightarrow B \rightsquigarrow C^{\text{phen}}.$$

Here the solid arrow indicates a directly supported intervention effect, the dashed arrows indicate mechanistic relations supported incompletely or indirectly, and the final wavy arrow indicates an inference from observable behavior to subjective state.

This notation summarizes the present evidential position:

$$\begin{aligned} M \longrightarrow D &\quad \text{strongly supported,} \\ D \dashrightarrow I &\quad \text{plausible and partly supported,} \\ I \dashrightarrow B &\quad \text{plausible but multiply mediated,} \\ B \rightsquigarrow C^{\text{phen}} &\quad \text{informative but non-equivalent.} \end{aligned}$$

These assignments are not permanent. A selective manipulation of dynamical stability or phase topology could strengthen the second arrow. Rich perturbational testing across graded behaviors could strengthen the third. Reliable no-report measures, postoperative reports, dream-like experience sampling, or other convergent indicators could improve inference across the fourth. The hierarchy is a map of current evidential strength, not a declaration that some transitions are forever inaccessible.

Necessary, Sufficient, and Constitutive Conditions

The hierarchy also prevents three forms of claim from being confused. A dynamical regime may be necessary for consciousness:

$$C^{\text{phen}} = 1 \quad \Rightarrow \quad D \in \mathcal{D}_{\mathsf{C}}.$$

It may be sufficient:

$$D \in \mathcal{D}_{\mathsf{C}} \quad \Rightarrow \quad C^{\text{phen}} = 1.$$

Or it may be constitutive, meaning that occupying the regime is part of what the relevant conscious organization consists in rather than merely a cause or prerequisite of it.

The anesthesia evidence does not establish any of these claims in full. Convergence across interventions is compatible with necessity, because a common dynamical disruption accompanies behavioral disconnection. But necessity would require showing that conscious experience never persists when the relevant regime is absent, and the dissociative case above makes this difficult to show.

Sufficiency would require restoring the regime and thereby restoring consciousness despite other impairments. No result reviewed here has done that. Constitutivity is stronger still, since it requires an account of why this form of organization should count as the relevant conscious process rather than as one of its enabling conditions.

The defensible claim is presently conditional:

$$D \in \mathcal{D}_{\text{viable}}$$

may be required for the form of integrated, environmentally responsive continuation characteristic of ordinary waking consciousness. That claim concerns a recognizable capacity. It does not identify the dynamical regime with subjective experience itself.

What "Global" Can and Cannot Mean

Descriptions of consciousness as a global regime can obscure as much as they clarify. "Global" need not mean that every cortical region participates equally, that the whole brain becomes synchronized, or that one scalar variable summarizes the state. It means that the relevant organization is not confined to one local representation and that activity in separated systems can enter mutually constraining continuations.

A global capacity may therefore be sparse, directional, and task-dependent. Some regions may coordinate only transiently. Some relations may be suppressive rather than excitatory. Some contents may remain local while others become widely available. The relevant structure may be represented by a time-dependent graph

$$\mathcal{G}_t = (V, E_t, W_t),$$

where nodes are neural populations, edges are effective relations, and weights encode direction, strength, phase dependence, or admissibility.

Loss of global capacity need not mean disappearance of every edge. It may mean that the graph no longer supports the paths needed to carry task-relevant distinctions:

$$\text{Path}_{\mathcal{T}}(i, j; t) = \varnothing$$

for critical population pairs, or that paths remain anatomically present but dynamically unusable.

This is why the bilateral increase in phase alignment discussed in Chapter 14 does not contradict loss of integration. More alignment along selected edges can coexist with loss of the differentiated, directional topology required for flexible continuation. Global integration is an organizational property, not the sum of pairwise synchronies.

The Hard Problem Remains Outside the Chain

Nothing in the hierarchy answers why an organized neural regime should be accompanied by experience. Even if every empirical arrow from molecular intervention through dynamics, integration, and

behavior were established, the result would remain a theory of the neural conditions and functional organization associated with consciousness:

$$M \longrightarrow D \longrightarrow I \longrightarrow B.$$

It would not derive phenomenal character from those relations.

The continuation-field framework may explain how a neural system selects content, preserves task-relative distinctions, binds distributed representations, repairs interrupted trajectories, and loses these capacities under anesthesia. These are substantial explanatory targets. None resolves why preserved continuation feels like anything from the inside.

This limitation should not be treated as an embarrassment requiring a speculative final step. A theory can explain the organization of conscious access and the conditions under which it fails without claiming to solve the metaphysical relation between organization and experience. Precision about the boundary of the theory is part of its explanatory value.

The Claim Part IV Can Defend

Part IV can defend the following sequence at different levels of confidence. Distinct anesthetic interventions alter neural dynamics. Some of these alterations converge across molecular mechanisms. The convergent changes concern stability and spatial phase organization rather than only the amount of activity. Those properties plausibly contribute to the system's capacity to preserve integrated continuation. Loss of that capacity plausibly contributes to behavioral unresponsiveness.

Part IV cannot yet defend the stronger sequence in which wave geometry has been shown to cause the loss of integration, the shared dynamical regime has been shown to be necessary and sufficient for consciousness, or behavioral unresponsiveness has been shown to imply absence of experience.

The distinction can be compressed into two statements:

$$\boxed{M \longrightarrow D \dashrightarrow I \dashrightarrow B}$$

is an empirically grounded research program, while

$$\boxed{D = C^{\text{phen}}}$$

is not a conclusion established by the evidence reviewed here.

The next chapter develops the strongest interpretation that remains available within these limits. Consciousness is treated provisionally as a preserved capacity for continuation: a regime in which distributed activity can remain mutually admissible, perturbations can be repaired, and task-relevant distinctions can be carried forward. The proposal concerns an enabling and organizing capacity. Whether that capacity is constitutive of consciousness or merely necessary for its ordinary waking form remains explicitly open.

Chapter 16

Consciousness as Preserved Capacity for Continuation

Chapter 15 drew the hierarchy connecting molecular intervention, altered dynamics, integrative capacity, behavior, and phenomenal consciousness with a deliberately mixed set of arrows, some solid, some dashed, one merely inferential. That hierarchy is a map of what has and has not been established. It is not yet a positive proposal. This chapter states the proposal the preceding three chapters license, at the strength they actually support, and no further.

The proposal is that consciousness, understood specifically as the capacity for integrated, environmentally responsive continuation characteristic of ordinary waking life, depends upon the nervous system occupying a dynamical regime in which activity in one region can serve as admissible context for activity elsewhere, in which perturbations can be absorbed and repaired rather than propagating into permanent derailment, and in which task-relevant distinctions can be carried forward across the interruptions that ordinary experience continually introduces. Anesthesia, on this account, does not straightforwardly delete the representations that make up a person's stored knowledge, memories, or perceptual capacities. It disrupts the transport relations, in the sense developed across Chapters 10 and 11, by which those representations become mutually available to one another and to the ongoing trajectory of behavior.

This framing has a specific consequence for where unconsciousness is located in the formal apparatus this book has built. The persistent disposition structure $\mathcal{S}$ is not what anesthesia principally targets. A person's capacities, memories, and learned associations remain materially encoded in synaptic organization that a general anesthetic does not erase, which is part of why a person wakes up as themselves rather than needing to relearn who they are. What anesthesia disrupts is the dynamics governing A_t, the admissibility field that determines which of the many contents $\mathcal{S}$ makes available are currently able to participate in a shared, unified trajectory. Contents remain present in the repertoire. What is lost is a sufficiently integrated path through A_t that would bind them into one continuing present.

It is worth being exact about what kind of claim this is, because the formal vocabulary makes it easy to overstate. This is a hypothesis about where in the system's overall organization anesthesia's effect is best located, stated in terms precise enough to be tested against the alternative Chapter 3 raised at the outset of this book: that ordinary recurrent-network dynamics, without any independently contributing wave geometry, are fully sufficient to produce the same destabilization and the same loss of integrative capacity that Chapters 13 and 14 documented. Nothing in this chapter, or in the anesthesia evidence reviewed so far, adjudicates between these two readings. Both are compatible with everything Part IV has shown.

The decisive experiment from Chapter 8 therefore returns here in its final form, applied now to the most global case this book considers. Two classes of intervention would be needed to separate the field-specific version of this proposal from its recurrent-network alternative. A geometry-selective intervention would alter wave phase, propagation direction, or spatial interference pattern during the transition into or out of an anesthetic state, while holding firing rates, spectral power, and other recurrent-state indicators approximately fixed, and would ask whether this selectively changes the holonomy defect $\delta_{\text{hol}}^{(\mathcal{T})}$ or the recovery score $R_{\mathcal{T}}$ introduced in Part III, applied here at the scale of global integrative capacity rather than a single content. A geometry-preserving intervention would do the reverse: alter recurrent connectivity or population-level dynamics directly, in a way expected to destabilize integration on the recurrent-network account, while holding wave geometry as fixed as the preparation allows, and would ask whether destabilization still occurs. If the first intervention changes integrative capacity while the second does not, the field-specific account gains real support. If the pattern runs the other way, the recurrent-network account is what the evidence favors, and this book's more elaborate apparatus would need to contract accordingly. Neither intervention has been performed in the anesthesia literature this part of the book has drawn on.

This leaves the chapter's central claim in a position that should by now be a familiar one to a reader who has followed the argument from Chapter 1 onward. Anesthesia provides some of the most striking evidence in this book that dynamics, not stored content, sit at the center of what integrated cognition requires, since a state indistinguishable in its representational repertoire from ordinary waking life can nonetheless fail to support any unified conscious continuation. That evidence supports treating consciousness, in its ordinary waking form, as a preserved capacity for continuation rather than as a fixed possession or a located structure. It does not yet support the stronger claim that wave interference specifically, rather than recurrent dynamics generally, is what does the preserving, and this chapter has tried to say so as plainly as Chapter 8 said it the first time, rather than letting four intervening chapters of increasingly specific evidence quietly promote a plausible hypothesis into an established one.

What can be said with more confidence is the shape of the claim itself, independent of which mechanism eventually turns out to fill it. Consciousness, on this account, is not a thing added to a sufficiently complex nervous system, and it is not straightforwardly identical to any single measured quantity, whether that quantity is drawn from spiking, from phase, or from field geometry. It is better understood as a capacity: the capacity of a system whose persistent structure supports an enormous combinatorial space of possible distinctions to continuously organize a coherent, admissible, and repairable path through that space, moment to moment, and to keep doing so across the interruptions, distractions, and perturbations that any embedded system operating in real time will inevitably encounter. Part III examined what happens when that capacity is tested locally, by a single distractor, against a single ongoing task. Part IV has examined what happens when the capacity itself is globally compromised, not by any local content failing to survive an interruption, but by the system losing, across the board, the dynamical conditions under which surviving an interruption is possible at all. Whether the compromised ingredient is best described in the vocabulary of recurrent circuits or in the vocabulary of traveling fields remains the open question this book has carried since Chapter 3 and has not resolved. That it is a dynamical, organizational capacity, rather than a fixed structure or a single localized substance, is the claim this part of the book can defend.

Part V now turns to consolidation rather than further argument. It gathers the formal definitions this book has developed, chapter by chapter, into a single compact reference, so that the vocabulary earned here through four parts of empirical argument can be used, checked, and extended without having to be rediscovered from the prose each time.

Part V

A Calculus of Continuation Fields

Chapter 17

Possibility Space and Admissibility Field

This chapter and the three that follow it gather what the preceding sixteen chapters earned, in compact and referenceable form. Nothing here is a new claim. Every definition restated in Part V was already stated once, in the place where the empirical or conceptual work justifying it was done, and this part exists only so that a reader who wants the vocabulary on its own, without the surrounding argument, has somewhere to find it. Where a definition's status remains open, that status is repeated here rather than quietly upgraded by the act of being collected into a tidier list.

The Persistent Disposition Space

$\mathcal{S}$ denotes the relatively persistent structure established by synaptic connectivity, learned representations, and other slowly changing parameters, first introduced in Chapter 2 to name one side of the underdetermination problem this book opened with. From $\mathcal{S}$, Chapter 5 built

$$P(\mathcal{S}),$$

the set of distinctions and continuations that structure can support under some attainable condition. Membership in $P(\mathcal{S})$ is dispositional, not actual: an item can belong to $P(\mathcal{S})$ while being neither currently accessible nor currently expressed. Chapter 1's argument that $\mathcal{S}$ alone does not fix a population's momentary behavior, and Chapter 3's insistence that a rival account can fill the resulting gap with ordinary recurrent state rather than any field-specific quantity, both remain exactly as they were. Collecting $P(\mathcal{S})$ here does not settle that dispute.

The Admissibility Field

$A_t \subseteq P(\mathcal{S})$ denotes the distinctions admissible at time t, playing the role Chapter 2 assigned to the generic fast-organizing variable q_t. Admissibility means more than physical possibility and less than actual expression: an admissible distinction is available to influence the current computation, enter working expression, constrain an action, or participate in a continuing trajectory. Chapter 5 was careful to keep A_t theory-neutral at the point of definition. Nothing in the notation A_t asserts that admissibility is implemented by a traveling wave. That further claim is the separate, still-open hypothesis

$$q_t \supseteq \mathcal{A}_t^{\text{wave}},$$

first named in Chapter 2 and never granted for free at any later point in the book, including here.

Expressed Content

$E_t \subseteq A_t$ denotes the distinctions presently expressed in measurable neural activity or behavior, with expression itself relativized to an observation map Ω wherever the scale of measurement matters, since Chapter 5 showed that a neural POP and a behavioral POP need not coincide. The three sets together give the inclusion chain that organizes the whole book's empirical vocabulary:

$$E_t \subseteq A_t \subseteq P(\mathcal{S}).$$

These levels are not interchangeable, and the book's recurring errors to avoid, named at various points from Chapter 5 onward, are precisely the errors of collapsing them: treating temporary absence from E_t as absence from A_t, or absence from A_t as absence from $P(\mathcal{S})$. Chapter 12's three senses of memory, dispositional, operative, and reconstructive, are this same distinction applied specifically to the question of what survives an interruption.

State and Trajectory

z_t denotes the instantaneous neural state, and

$$\gamma_{[t_0,t_1]} = \{z_t : t \in [t_0, t_1]\}$$

denotes the trajectory traced over an interval, a distinction Chapter 2 introduced specifically because Part III would later need to compare a pre-distraction trajectory, a post-distraction trajectory, and a counterfactual uninterrupted trajectory without conflating an instant with an interval. The dynamics governing z_t were written in Chapter 2 as

$$\dot{z}_t = F_{\mathcal{S}}(z_t, u_t; \mathcal{A}_t^{\text{wave}}),$$

with the explicit caveat that this form is only licensed once the field hypothesis is granted; the theory-neutral version, in which the admissibility contribution is denoted by the generic q_t rather than the field-specific $\mathcal{A}_t^{\text{wave}}$, remains the form this book is actually entitled to at any point where the field hypothesis has not been separately earned.

What This Chapter Does Not Add

Nothing above resolves the underdetermination problem of Part I, decides between the field account and the recurrent-network competitor, or upgrades any tier-1 finding to tier 2 or tier 3. The value of gathering $\mathcal{S}$, $P(\mathcal{S})$, A_t, E_t, z_t, and γ into one place is purely organizational: a reader returning to this book for its formal apparatus, rather than for its argument, should be able to find the apparatus here without having to reconstruct it from sixteen chapters of prose. The next chapter does the same for the four operators built on top of these sets.

Chapter 18

The Operators: POP, REFUSE, BIND, COLLAPSE

Chapter 5 defined four transformations on the sets gathered in the previous chapter, each with a side-condition specifying what would have to fail for the operator not to apply. This chapter restates the four operators in compact form, together with the falsification conditions Chapters 5 through 11 attached to each. Where the tier evidence for a given operator was assessed in a later chapter, that assessment is noted rather than repeated in full.

POP: Entry Into Expression

$$\mathrm{POP}_{t,\Delta t}(x) : x \in A_t \setminus E_t \longmapsto x \in E_{t+\Delta t}.$$

Prior admissibility is required; an event that forces a wholly unavailable distinction into the system by changing $\mathcal{S}$ itself is learning or structural modification, not POP. Expression is relativized to an observation map Ω, so a neural POP need not coincide with a behavioral one. A candidate marker M_t realizes POP only if it predicts, and ideally causes, the specific transition of x from admissibility into expression, not merely correlating in time with some transition or other. Chapter 4 identified gamma bursting as a plausible tier-1 marker of content entry in the Lundqvist working-memory paradigm; Chapter 6 then showed why spatial prediction alone cannot promote an operator-marker correspondence to tier 2.

REFUSE: Active Exclusion From Admissibility

$$x \in P(\mathcal{S}), \qquad x \notin A_t,$$

with the counterfactual requirement that under a matched context C_t^{-r} lacking the candidate refusing process r, x would have had non-negligible probability of becoming admissible:

$$\rho_r(x,t) = \Pr(x \in A_{t+\Delta t} \mid C_t^{-r}) - \Pr(x \in A_{t+\Delta t} \mid C_t) > \varepsilon.$$

Without this condition, REFUSE would name the generic condition of not happening rather than a specific operation, since nearly every distinction in $P(\mathcal{S})$ is absent from expression at any given moment. REFUSE is distinct from erasure: x remains in $P(\mathcal{S})$, only excluded from A_t, and is in principle recoverable without being rebuilt. Chapter 6 found correct-versus-error trial contrasts insufficient to establish ρ_r causally, since such contrasts are observational rather than interventional; the required intervention has not been performed in the literature this book draws on.

BIND: Formation of a Jointly Operative Distinction

$$\mathrm{BIND}_{\kappa}(x, y) : (x, y) \longmapsto \langle x, y \rangle_{\kappa},$$

with the joint distinction required to carry task-relevant information not exhausted by its components, formalized through the partial information decomposition

$$I(X, Y; T) = R + U_X + U_Y + \mathrm{Syn}(X, Y; T),$$

where $\mathrm{Syn}(X, Y; T) > 0$ is the operative criterion. Temporal coincidence, anatomical proximity, and phase synchronization are candidate mechanisms of binding, never definitions of it. Chapter 6 found the spatial-computing literature to supply a plausible substrate for BIND without demonstrating it under this criterion, since the relevant partial information decomposition has not been performed on that data; this was treated as the calculus doing its discriminatory work rather than as a shortfall in the theory.

COLLAPSE: Resolution Under Constraint

$$\mathrm{COLLAPSE}_{t,\Delta t} : \Gamma(A_t) \longmapsto \gamma^{*}_{[t,t+\Delta t]},$$

the resolution of a family of admissible continuations into one realized trajectory, compatible with either deterministic or stochastic dynamics. COLLAPSE is not continuation failure, rupture, or disintegration, which name the separate case in which the system fails to sustain any organized resolution at all; keeping these apart was necessary once Part III and Part IV both needed to describe failure without overloading a single word to mean two different things. A candidate marker of COLLAPSE must predict a measurable reduction in the conditional entropy of future trajectories, not merely an increase in synchrony that looks organized without being shown to mark the specific point at which alternatives ceased to remain available.

Composition and the Three-Tier Test, Restated

The four operators compose without a mandatory order; REFUSE, POP, BIND, and COLLAPSE were illustrated in Chapter 5 as one possible sequence within a working-memory trial, not as a fixed pipeline. Each operator carries its own instance of the book's recurring tier test: tier 1 requires that a wave property be informative about the operator's occurrence, tier 2 requires that intervention on the wave property change the operator's probability or timing, and tier 3 requires that this effect survive conditioning on the strongest available recurrent-state account. No operator examined in this book has been shown to clear tier 3. REFUSE and BIND remain at tier 1 by the assessments of Chapter 6; POP's evidence is comparably placed; COLLAPSE's clearest empirical treatment came in Part III, where the resolution of post-perturbation continuations was reframed as task-relative recovery rather than assessed as a freestanding operator in isolation, a reframing the next chapter restates in compact form. Continuation failure, the loss of an organized regime altogether, remains distinct from COLLAPSE throughout.

Chapter 19

Transport, Holonomy, and Task-Relative Equivalence

Part III replaced literal return with a task-relative notion of recovery and then asked what it would take to describe the transport of a representation through an interruption, rather than only comparing its endpoints. This chapter gathers both pieces: the equivalence relation from Chapter 10 and the transport apparatus from Chapter 11, restated compactly and cross-referenced to where each was earned.

Task-Relative Equivalence

Given a task map $\pi_{\mathcal{T}} : Z \to Y_{\mathcal{T}}$, specified independently of the outcome it is later used to explain,

$$z_1 \sim_{\mathcal{T}} z_2 \quad \Longleftrightarrow \quad \pi_{\mathcal{T}}(z_1) = \pi_{\mathcal{T}}(z_2).$$

This is the replacement for pointwise return: two neural states can be task-equivalent while being physically distinct, and successful recovery requires only $z^+ \sim_{\mathcal{T}} z^-$, not $z^+ = z^-$. The approximate tolerance version $\approx_{\mathcal{T},\delta}$ is not transitive, which matters once equivalence is asked to hold along an extended trajectory rather than at a single comparison. Equivalence is always indexed to a task; the same transport can be harmless relative to a coarse probe and destructive relative to a finer one, formally $z^+ \sim_{\mathcal{T}_1} z^-$ while $z^+ \not\sim_{\mathcal{T}_2} z^-$.

The Recovery Score

$$R_{\mathcal{T}} = 1 - \min\left\{1, \frac{D_{\mathcal{T}}(\gamma^d, \widehat{\gamma}^0)}{\Lambda_{\mathcal{T}}}\right\},$$

comparing the observed post-perturbation trajectory against a counterfactual uninterrupted one, the latter never directly observable and requiring its own validated estimator. $R_{\mathcal{T}}$ is graded rather than binary, and its normalization $\Lambda_{\mathcal{T}}$ must be sourced explicitly, since a scale drawn from correct-versus-error separation and one drawn from undistracted-trial variability answer different questions and are not interchangeable.

Transport and the Connection

Where the task base $M_{\mathcal{T}}$, the realization fibers $F_{m,y}$, the transformation group G, and the task-preserving subgroup $K_{\mathcal{T}}$ can all be independently specified, a connection $\omega^{(\mathcal{T})}$ supplies a rule for comparing representational frames across contexts, estimated from data the recovery question does not itself depend on. Transport along a path is

$$U_{\Gamma}^{(\mathcal{T})} = \mathcal{P} \exp\left(- \int_{\Gamma} \omega^{(\mathcal{T})} \right),$$

and the holonomy around the comparison loop formed by the distracted path and the reverse of the counterfactual uninterrupted path is

$$H_{C}^{(\mathcal{T})} = \mathcal{P} \exp\left(- \oint_{C} \omega^{(\mathcal{T})} \right).$$

The Gauge of Persistence

The criterion for successful persistence is

$$H_{C}^{(\mathcal{T})} \in K_{\mathcal{T}},$$

not $H_{C}^{(\mathcal{T})} = I$. This is the chapter's single most important piece of vocabulary, restated here because it is the point most liable to be flattened back into a literal-return standard if quoted out of context. Multiple neural realizations represent the same continuing thought exactly when they differ only by transformations lying in $K_{\mathcal{T}}$, transformations that preserve every future discrimination the task in question requires. A residual transformation outside $K_{\mathcal{T}}$, however small in the ambient metric of G, constitutes failure if it crosses a task-relevant boundary; a large transformation confined to $K_{\mathcal{T}}$ is cognitively harmless.

Chapter 11 was explicit that this entire apparatus is conditional on empirical support that has not yet been produced: on whether the fibers exhibit the local regularity a bundle description requires, on whether the connection generalizes to held-out trials and novel perturbation orderings, and on whether the physical rotating wave documented in Part III's source literature can be shown, rather than merely suggested, to determine the connection rather than being a downstream signature of ordinary recurrent recovery. None of those conditions has been met. Gathering the formalism here does not meet them either; it only makes the formalism available for whichever future evidence eventually does.

Chapter 20

The Full Chain and Its Open Arrow

The three preceding chapters gathered the book's vocabulary in pieces: the sets in Chapter 17, the operators acting on them in Chapter 18, the transport structure describing their persistence through interruption in Chapter 19. This final chapter of Part V puts the pieces together into a single chain and states, as compactly as the material allows, which links in that chain are load-bearing and which remain open.

The Chain

A single linear arrow, of the form $\mathcal{S} \to A_t \to E_t \to \gamma_t$ with one operator label per arrow, would misrepresent what Chapter 5 actually defined. REFUSE acts within A_t itself, excluding a distinction rather than carrying A_t into E_t; BIND can act on distinctions within A_t or already within E_t; and COLLAPSE maps the family of admissible continuations $\Gamma(A_t)$ to a realized trajectory, not E_t directly to one. The faithful consolidation is therefore not a single pipeline but a set of coordinated transformations on the structured state $\Xi_t = (P(\mathcal{S}), A_t, E_t, z_t, \Gamma_t)$ introduced in Chapter 5:

$$\mathcal{S} \longmapsto P(\mathcal{S}), \qquad A_t \subseteq P(\mathcal{S}), \qquad E_t \subseteq A_t,$$

$$\mathrm{REFUSE}_t : A_t \longmapsto A_t \setminus \{x\}, \qquad \mathrm{POP}_t : A_t \setminus E_t \longmapsto E_{t+\Delta t},$$

$$\mathrm{BIND}_\kappa : (x, y) \longmapsto \langle x, y \rangle_\kappa, \qquad \mathrm{COLLAPSE}_{t,\Delta t} : \Gamma(A_t) \longmapsto \gamma^*_{[t,t+\Delta t]}.$$

The persistent disposition space $\mathcal{S}$ supplies the possibility space $P(\mathcal{S})$; the admissibility field A_t, whose relation to a structured wave geometry is the field hypothesis rather than a settled fact, narrows that space to what is currently live; REFUSE, POP, and BIND act as needed on A_t and E_t without a mandatory order, exactly as Chapter 5 illustrated with its working-memory example; COLLAPSE resolves the resulting family of admissible continuations, not the expressed set itself, into a realized trajectory. Every transformation in this summary was earned in a specific earlier chapter and carries that chapter's qualifications with it. None is asserted here for the first time, and none is a fixed stage in a linear sequence the others must wait on.

Superimposed on this structure, Part III's transport apparatus describes what happens to it across an interruption. Recovery is not $z^+ = z^-$ but $z^+ \sim_{\mathcal{T}} z^-$, and where a connection can be independently estimated, persistence through the interruption is $H_C^{(\mathcal{T})} \in K_{\mathcal{T}}$ rather than $H_C^{(\mathcal{T})} = I$. This is the same structure, examined under perturbation rather than during undisturbed operation.

The Chain at Global Scale

Part IV's hierarchy is the same structure applied at a different scope. Where Chapters 1 through 12 asked how a single content is selected, expressed, bound, and recovered, Chapters 13 through 16 asked whether the system retains the capacity to do any of this at all:

$$M \longrightarrow D \dashrightarrow I \dashrightarrow B \rightsquigarrow C^{\text{phen}}.$$

The solid arrow, $M \to D$, is the best-supported link in this second chain, matching the strength of evidence for the corresponding tier-1 claims earlier in the book. The dashed arrows, $D \dashrightarrow I$ and $I \dashrightarrow B$, are plausible and partly supported, corresponding to tier-1 and fragmentary tier-2 evidence rather than to anything stronger. The final arrow is not dashed but wavy, $B \rightsquigarrow C^{\text{phen}}$, marking an inferential bridge rather than a causal claim of the same kind as the others. This is the open arrow named in this chapter's title, and it is open for a reason internal to the subject matter rather than for want of better experiments: no measurement of dynamics, integration, or behavior settles, by itself, whether the organization those measurements track is accompanied by experience.

What Is Established and What Is Not

Collecting the chain into one diagram invites a temptation this book has tried to resist throughout, which is to let the tidiness of the picture stand in for the strength of the evidence behind it. It does not. The arrows $\mathcal{S} \to A_t$ and $A_t \to E_t$ are underwritten by a real and precisely stated underdetermination problem, not by a demonstration that wave geometry specifically fills the admissibility role; Chapter 3's recurrent-network competitor remains, at the close of this book, exactly as live as it was when first introduced. The operator arrows are underwritten by tier-1 evidence and, for a subset of cases reviewed in Part II, plausible but unconfirmed tier-2 mechanisms. The transport arrows are underwritten by a formally careful apparatus whose empirical conditions, listed in full in Chapter 11, have not yet been met by any experiment this book has reviewed. The global-scale chain is underwritten by genuine convergence across mechanistically distinct anesthetic agents, an inference this book has called triangulation rather than proof, and its final arrow is, by the nature of the question it asks, not the kind of thing further convergence alone could close.

What the Chain Is For

The value of this diagram is not that it proves the continuation-field hypothesis. Nothing in this book has proved it, and the decisive experiment named in Chapter 8 and reapplied in Chapters 11 and 16 has not, at any point, been performed. The value is that the diagram states precisely what such an experiment would need to show, and where in the chain it would need to show it, for the hypothesis to move from a coherent and evidentially serious proposal to a demonstrated one. A theory that cannot say this much about itself is not yet in a position to be tested. A theory that says it too vaguely invites its own evidence to be overread. This book has tried, chapter by chapter, to avoid both failures. Whether it has succeeded is a question for the reader rather than for its author, and the Afterword that follows closes by returning to exactly that question one final time.

Afterword: What Remains Theoretical

A book built around a hypothesis that has not been established owes its reader, at the end, the same discipline it tried to maintain throughout: a plain statement of what is now known that was not known at the start, and what remains exactly as open as it was in the Preface.

What is established, at this point, does not depend on the continuation-field hypothesis being correct. Mixed selectivity gives the prefrontal cortex a combinatorial representational capacity that a population of cleanly tuned neurons could not match [14], and this capacity creates a genuine problem: something fast enough to track a train of thought must select, moment to moment, which of a neuron's many possible functional roles is currently operative. Alpha and beta rhythms are associated, in specific and well-controlled paradigms, with task rules and goals, while gamma-range activity is associated with the moment-to-moment expression of held content [9], though this association has not been shown to hold as a general law across every area, task, and species. Cortical oscillatory power forms spatially organized patches whose geometry shifts with task demands and predicts trial-by-trial behavioral success [10, 5]. Extracellular electric fields can influence the spiking of nearby neurons through ephaptic coupling, a real biophysical possibility demonstrated under multiple conditions [12, 13], although its magnitude, selectivity, and computational importance in ordinary behaving cortex remain unresolved. A rotating traveling wave follows distraction, and the completeness of that rotation predicts whether a task-relevant train of thought is successfully continued [3]. Three anesthetic agents acting on three distinct primary molecular targets converge on a shared destabilization of cortical dynamics and a shared reorganization of phase alignment as behavioral responsiveness is lost [7, 2], a convergence that does not imply their downstream circuit consequences are unrelated, only that no single receptor-level account is sufficient on its own. Each of these findings would remain true, and remain worth knowing, even if every further claim this book has built on top of them turned out to be wrong.

What remains open is, in one sense, a single question asked at increasing scales. Chapter 8 asked it first, in the narrowest setting this book examined: whether the specific geometry of wave interference, as opposed to the mere fact that oscillatory activity correlates with cognition, performs a computation not reducible to a description of spiking and recurrent synaptic dynamics. Chapter 11 asked the same question of a rotating wave and a recovered train of thought. Chapter 16 asked it of a destabilized cortex and a lost capacity for unified experience. In every case, the answer given by the evidence reviewed in this book is the same: plausible, worth taking seriously, and not yet shown. The recurrent-network account introduced in Chapter 3 to stand as this book's most serious competitor has not been refuted at any point along the way, and a reader who finds that account more parsimonious than the one this book has developed has not misread the evidence. They have read it correctly and weighted its remaining uncertainty differently than this book has chosen to.

The claim this book can defend, stated at the strength it has actually earned rather than the strength that would make for a more satisfying conclusion, is this: cognition may require a continuously reconstructed field of admissibility, and the geometry of that field, if it exists as an independent causal contributor rather than as a byproduct of recurrent computation, would determine which of a fixed representational repertoire's latent distinctions can become a coherent continuation of thought right now. This is a stronger claim than the popular observation that brain waves organize thought, since it specifies a mechanism, a formal criterion for testing it, and a named competitor it has to outperform to be believed. It is a substantially weaker claim than the suggestion that wave dynamics have been shown to create consciousness, since nothing in this book has shown anything of the kind, and Chapter 15 went to some length to explain why the evidential chain connecting neural dynamics to phenomenal experience contains an inferential bridge rather than a causal arrow at its final link, a link no amount of further convergence at the dynamical level alone could close.

One experiment, stated in the same form each time it recurred, would do more than anything else to move this book's central hypothesis from a coherent and evidentially serious proposal toward a demonstrated one. It requires perturbing wave geometry directly, its spatial phase, its direction and speed of propagation, the completeness of a rotational pattern, or the alignment between interacting frequency bands, while holding fixed everything a recurrent-network account would need in order to explain the same outcome without appeal to that geometry at all: firing rate, spectral power taken independently of spatial or phase structure, sensory input, and task difficulty. If a decoded operator, a recovery score, or a measure of integrative capacity tracks the manipulated geometry specifically, changing when and only when the geometry changes, this book's central hypothesis gains the kind of support nothing reviewed here has yet provided. If it does not, the theory this book has built should contract, cleanly and without resistance, to the more modest claim that oscillatory dynamics are informative and plausibly causally entangled with cognition, without being an independent computational resource in their own right.

That experiment has not been performed. This book ends, as it must, without knowing which way it would come out.

Bibliography

[1] Anastassiou, C. A., Perin, R., Markram, H., & Koch, C. (2011). Ephaptic coupling of cortical neurons. *Nature Neuroscience*, 14(2), 217–223.

[2] Bardon, A. G., Ballesteros, J. J., Brincat, S. L., Roy, J. E., Mahnke, M. K., Ishizawa, Y., Brown, E. N., & Miller, E. K. (2025). Convergent effects of different anesthetics on changes in phase alignment of cortical oscillations. *Cell Reports*, 44, 115685.

[3] Batabyal, T., Brincat, S. L., Donoghue, J. A., Lundqvist, M., Mahnke, M. K., & Miller, E. K. (2025). State-space trajectories and traveling waves following distraction. *Journal of Cognitive Neuroscience*.

[4] Berger, H. (1929). Über das Elektrenkephalogramm des Menschen. *Archiv für Psychiatrie und Nervenkrankheiten*, 87, 527–570.

[5] Chen, Z., Brincat, S. L., Lundqvist, M., Loonis, R. F., Warden, M. R., & Miller, E. K. (2026). Oscillatory control of cortical space as a computational dimension. *Current Biology*, 36(2), 402–414.e5.

[6] Eisen, A. J., Kozachkov, L., Bastos, A. M., Donoghue, J. A., Mahnke, M. K., Brincat, S. L., Chandra, S., Tauber, J., Brown, E. N., Fiete, I., & Miller, E. K. (2024). Propofol anesthesia destabilizes neural dynamics across cortex. *Neuron*, 112(16), 2799–2813.e9.

[7] Eisen, A. J., Bardon, A. G., Ballesteros, J. J., Bastos, A. M., Donoghue, J. A., Mahnke, M. K., Brincat, S. L., Roy, J. E., Ishizawa, Y., Brown, E. N., Fiete, I., & Miller, E. K. (2026). Similar destabilization of neural dynamics under different general anesthetics. *Cell Reports*, 45, 117048.

[8] Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, 19(4), 1273–1302.

[9] Lundqvist, M., Rose, J., Herman, P., Brincat, S. L., Buschman, T. J., & Miller, E. K. (2016). Gamma and beta bursts underlie working memory. *Neuron*, 90(1), 152–164.

[10] Lundqvist, M., Brincat, S. L., Rose, J., Warden, M. R., Buschman, T. J., Miller, E. K., & Herman, P. (2023). Working memory control dynamics follow principles of spatial computing. *Nature Communications*, 14(1), 1429.

[11] Miller, E. K., Brincat, S. L., & Roy, J. E. (2026). Analog cognition and consciousness. *The Journal of Neuroscience*, 46(33), e0711262026.

[12] Pinotsis, D. A., & Miller, E. K. (2023). In vivo ephaptic coupling allows memory network formation. *Cerebral Cortex*, 33(19), 10353–10365.

[13] Pinotsis, D. A., & Miller, E. K. (2026). Ephaptic coupling can explain variability in neural activity. *Cerebral Cortex*, 36(6), bhag098.

[14] Rigotti, M., Barak, O., Warden, M. R., Wang, X.-J., Daw, N. D., Miller, E. K., & Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature*, 497(7451), 585–590.

[15] Williams, P. L., & Beer, R. D. (2010). Nonnegative decomposition of multivariate information. *arXiv preprint* arXiv:1004.2515.